Decoding the real-time neurobiological properties of incremental semantic interpretation.

Hun S. Choi, William Marslen-Wilson, Bingjiang Lyu, Billi Randall & Lorraine K. Tyler

Centre for Speech, Language and the Brain, Department of Psychology, University of Cambridge

Abstract

Communication through spoken language is a central human capacity, involving a wide range of complex computations that incrementally interpret each word into meaningful sentences. However, surprisingly little is known about the spatiotemporal properties of the complex neurobiological systems that support these dynamic predictive and integrative computations. Here, we focus on prediction, a core incremental processing operation guiding the interpretation of each upcoming word with respect to its preceding context. To investigate the neurobiological basis of how semantic constraints change and evolve as each word in a sentence accumulates over time, in a spoken sentence comprehension study we analysed the multivariate patterns of neural activity recorded by source-localised electro/magnetoencephalography (EMEG), using computational models capturing semantic constraints derived from the prior context on each upcoming word. Our results provide insights into predictive operations subserved by different regions within a bi-hemispheric system which over time, generate, refine and evaluate constraints on upcoming words.

Keywords: Bayesian language modelling, incremental prediction, semantics, representational similarity analysis, electro/magnetoencephalography

Introduction

Spoken language comprehension involves a variety of rapid computations that transform the auditory input into a meaningful interpretation. When listening to speech, our primary percept is not of the acoustic-phonetic detail, but of the speaker's intended meaning. This effortless transition occurs on millisecond timescales, with remarkable speed and accuracy, and without any awareness of the complex computations on which it depends. How is this achieved? What are the processes and representations that support the transition from sound to meaning, and what are the neurobiological systems in which they are instantiated?

Understanding the meaning of spoken language requires listeners to access the meaning of each word that they hear and integrate it into the ongoing semantic representation in order to

incrementally construct a syntactically-licensed semantic representation of the sentence (Tyler and Marslen-Wilson 1977; Marslen-Wilson and Tyler 1980; Kamide et al. 2003; Hagoort et al. 2009). Research to date provides a broad outline of the neurobiological language system and of the variables involved in language comprehension (Hickok and Poeppel 2007; Marslen-Wilson and Tyler 2007; Friederici 2011; Kutas and Federmeier 2011; Price 2012; Bornkessel-Schlesewsky and Schlesewsky 2013; Hagoort 2013; Matchin and Hickok 2019), but surprisingly little is known about the specific spatio-temporal patterning and the neurocomputational properties of the incremental processing operations that underpin the dynamic transitions from the speech input to the meaningful interpretation of an utterance.

This is our goal in the present study where we probe directly the dynamic patterns of time-sensitive neural activity that are elicited by spoken words, focusing on the semantic constraints they generate on upcoming words, and the incremental processes that combine them into semantically coherent utterance interpretations. We use computational linguistic analyses of language corpora to build quantifiable models of semantic constraint and mismatch, where the latter reflects the processing demands of interpreting the upcoming word given the properties of prior constraints (Hale 2001; Levy 2008). Based on these cognitive models, we employ Representational Similarity Analysis (RSA) to probe the different types of neural computation that support dynamic processes of incremental interpretation, using source-localised MEG + EEG (EMEG) imaging to capture the real-time electrophysiological activity of the brain. RSA enables us to compare the (dis)similarity structure of our theoretically relevant models with the (dis)similarity structure of observed patterns of brain activity, revealing how different information types are encoded in different brain areas over time.

In a previous EMEG study, involving single spoken words, we used these methods to map out the spatio-temporal dynamics of the word recognition process (Kocagoncu et al. 2017). Using RSA to test quantifiable cognitive models of key analysis processes as they occur in real time in the brain, we identified the cortical regions that support the early phonological and semantic competition between cohort candidates as a word is heard, and the dynamic process of convergence on a single candidate and its unique semantic representation as the uniqueness-point (UP) approaches (i.e. the point at which the word can be differentiated from its word-initial cohort and is uniquely recognisable (Marslen-Wilson 1987)). Hence, identifying the UP plays an important role in interpreting the timing of linguistic processing with respect to the input word. In a subsequent study, placing spoken words in a minimal phrasal context (e.g., *yellow banana*), we constructed RSA models of the semantic constraints generated by the adjective (*yellow*) to determine how these interacted with the processing of the following noun (*banana*). Consistent with previous behavioural and ERP results (Marslen-Wilson 1975; Kamide et al. 2003; DeLong et al. 2005; Bicknell et al. 2010), we found

early effects of prior probabilistic semantic constraints on lexical processing (within 150-200ms of word onset), where the timing of these effects reflects the prior access of potential word candidates driven by the sensory input (Klimovich-Gray et al. 2019). These studies suggest an underpinning lexical access process where lexical contents can be made available very soon after word onset for interaction with contextual constraints.

In the context of these two studies, the current study aims to determine how the rich contextual constraints incrementally combine words into a meaning interpretation and how this interpretation modulates the processing of subsequent words in the utterance. Critical to this study is the development of the appropriate quantifiable measures of the relevant properties of the sentential processing environment, as the basis for the RSA models used to probe the real-time brain activity elicited by hearing the test sentences.

Within the broad context of predictive processing frameworks (Kuperberg and Jaeger 2016), we investigated the role of semantic constraint elicited by the incrementally developing context in sentences such as "*The experienced walker chose the path*", including its subject, verb and object, in generating a message-level interpretation. To do this we used language models of constraint and mismatch derived by combining the behavioural responses from sentence completion studies with the latent Dirichlet allocation (LDA) approach of topic modelling (Griffiths and Steyvers 2004). These models were used to construct RSA models of semantic constraints, as they evolve over a spoken utterance, and to look at the spatiotemporal pattern of model fit for each processing dimension being tested (Kocagoncu et al. 2017). Importantly, the cognitive models that test for effects of semantic constraints and their integration into the developing sentence are probabilistic and experiential in nature, reflecting language as people experience it in the real world and providing the type of quantifiable data necessary to calculate rich multivariate representational models. This avoids the limitation of relying on categorical distinctions between stimuli which fail to capture the multifaceted richness of linguistic representations and the probabilistic nature of language.

Our primary interest here is in what we call "combined constraints" on upcoming words, the cumulative constraints generated by the set of words comprising the prior context. In this study, we developed a set of contextual constraint models in order to illuminate the temporal progression of predictive processing as each word (i.e. verb and complement noun (CN)) incrementally unfolds over time. This enables us to illustrate the spatiotemporal dynamics of the cumulative effects of constraints and to determine how far these constraints are neurally expressed.

In common with recent accounts of incremental processing of speech inputs, we expect to see the computation of constraints as each word is being recognised (Marslen-Wilson 1975; Marslen-Wilson

and Tyler 1980; DeLong et al. 2014). The RSA models, as described above, primarily focus on modelling these constraints and the relative timing with which they appear as the utterance unfolds over time. We also investigate the mismatch effect between the context and a target word (CN) that captures the difficulty of semantically processing the target word with respect to the constraint imposed by the prior context, based on its semantic properties. Together the timing and location of the effects captured by these models reveal a picture of when and where the human brain activates and utilizes constraints at the semantic level.

Overview

To determine the spatiotemporal neural properties of incremental semantic interpretation during language comprehension, we developed models of the incremental constraints that the context imposes on the meanings of upcoming words, and the mismatch between an upcoming word and its fit into the prior context. We tested these models against the spatiotemporal properties of the source-localised EMEG data to compare the similarity structure of our theoretically relevant models. We tested for the timing of the model fit generated for these models at different time points within a language mask that includes a set of brain regions comprising a bilateral fronto-temporo-parietal language system which has been frequently reported in the literature (Binder et al. 2009). We asked when and where each of our key models – of semantic constraint, and mismatch - would fit the brain data; when and where is there an effect of the subject noun phrase (SNP) semantic constraint, how does it change as a subsequent verb is processed, and what is the scope of these constraint effects on upcoming words?

In order to model incrementally developing constraint over time, we obtained measures of semantic prediction at two different points in a sentence – immediately after the SNP [“the experienced walker”] and after the combination of the SNP+verb [“the experienced walker chose…”]. In this way, we aimed to characterize the changing patterns of prediction as a verb is combined with the initial SNP context. To do this, we conducted two separate behavioural studies with different participants in which they were asked to complete a sentence either after hearing the SNP fragments (study 1) or after hearing the SNP+verb fragments (study 2). We then extracted main verbs from the first behavioural study and CNs from the second behavioural study, allowing us to infer the predictive state of the brain throughout the sentence.

However, in natural speech comprehension, prior constraints are relatively broad, so that specific words are rarely strongly predicted (Luke and Christianson 2016). Particularly, during the early stage of sentence processing, the context (SNP or SNP+verb) rarely provides a strong prediction of a

particular upcoming word, leading to high uncertainty (entropy) in word-level constraints (Kuperberg 2016). Therefore, we applied topic modelling to each unique word provided by participants in the behavioural studies, in order to characterize constraints derived from the rich semantic (topic) representation associated with each unique word in a Bayesian framework of incremental predictive processing. To model prediction at a more abstracted semantic level, we combined the topic distributions of the continuation data into semantic "blends" of word candidates, modelling the conditional probability distribution $P(topic|full\ context)$ (see 3.3 in Methods). Then, we computed entropy (see 3.4 in Methods) of the blend to quantify the overall constraint strength which was tested against the EMEG data during relevant epochs as described in Table 1 (see also Figure 1), in order to investigate the incremental development of semantic constraint. Finally, in order to investigate how the constrained words are evaluated and incorporated into the prior context (SNP+verb), we also characterized the EMEG data using a pattern of mismatch between the predicted and the target semantics (see 3.5 in Methods).

In light of the claims that semantics is represented bilaterally (Price 2010, 2012; Wright et al. 2012), our approach provides an opportunity to determine whether different kinds of semantic computations are represented differentially across the hemispheres. We expected the predictive computations based on this information to involve bilateral anterior temporal and frontal areas with the right hemisphere involved in construction of a broader semantic representation and the engagement of the context (Beeman and Chiarello 1998; St George et al. 1999; Seger et al. 2000; Jung-Beeman 2005).

Methods

1) Participants. Fifteen participants (7 females; average age: 24 years; range: 18-35 years) took part in the study. They were all native British English speakers and right-handed with normal hearing. Two participants were excluded from the analysis, one because of sleepiness during the EMEG study and the other because of poor quality EEG recordings. Informed consent was obtained from all participants and the study was approved by the Cambridge Psychology Research Ethics Committee.

2) Stimuli: We constructed 200 spoken sentences consisting of a SNP (e.g. "*the experienced walker*"), followed by a verb (e.g. "*chose*") which in turn was followed by a CN (e.g. "*path*"). The sentence sets were constructed in the following way. First, we chose verbs from the VALEX database (Korhonen et al. 2006) that occurred with (at least) two different complement structures, one was a simple transitive direct object (DO) structure (e.g. "… *chose the path*…"), and the other was one of three

other possible complement structures including sentential complement (SC; "… *denied that the court* …), infinitival complement (INF; "… *wanted to become …"),* and prepositional phrase complement (PP; "… *fled to the forest …").* For 72% of the stimuli, the DO complement structure was more frequent (according to the subcategorization frame (SCF) information in VALEX; (Korhonen et al. 2006)) with the average probability of 0.499±0.12 (mean ± sd). By adding some variability to the function words of the complement phrase, we aimed to improve the generalizability of our results to any natural spoken sentence with varying subcategorization structures.

To ensure variability in the predictability of the CNs, we varied the probability of these nouns with the preceding verb and the complement function word according to Google Books n-gram frequencies. Note that this variability was controlled when running the analysis by including the frequency of a word to which the epoch was aligned to as one of the covariates and partialling out when correlating the data and model representational dissimilarity matrices (e.g. SN frequency at epoch 1, verb frequency at epoch 2 and CN (content word) frequency at epoch 3). This process resulted in 200 sentences with four repetitions of the SNP + verb combination (see Figure 2), consisting of varying complement structures (i.e. DO, SC, INF and PP) with different complement content words. This ensured sufficient variability between trials in the ease with which the content word in the complement could be integrated into the ongoing sentential representation, given the constraints provided by the preceding context. Just as for the lexical frequency, we controlled for the repetition effect of the SNP + verb combination by including it as another covariate. In summary, we partialled out the effects of 1) lexical frequency of a word to which an epoch is aligned and 2) repetition of stimuli across trials.

The sentences were spoken by a native female British English speaker and were recorded in a soundproof booth. In the experiment, participants were asked to listen to these sentences attentively while we recorded their brain activity using EMEG. There was no explicit task for them to perform since tasks are known to invoke domain general brain systems over and above any domain-specific language effects (Campbell and Tyler 2018). All stimuli were pseudo-randomized and counter-balanced across participants. We followed the standard procedure for presenting auditory stimuli as in our previous studies (Kocagoncu et al. 2017; Klimovich-Gray et al. 2019).

3) Incremental models of predictive processing

In this study, we focused on the two different incremental computations: 1) constraint and 2) evaluation in order to investigate the neurobiological underpinnings of how the preceding context guides the interpretation of an upcoming word. To do this, we combined behavioural data with computational models of semantics as described below.

3.1) Behavioural studies

To model incrementally evolving constraints over the SNP, verb and CN, we conducted two separate behavioural studies. In the first experiment, 24 participants (who did not take part in the main experiment or the second behavioural study) heard each unique SNP (e.g. "*The experienced walker* …") and provided a sentence continuation after the SNP (e.g. "… hiked through the mountains", "… chose a less travelled path" etc). We extracted the main verb from each sentence continuation and used these data with topic representations (see 3.2 in Methods) to capture predicted verb semantics. In the second experiment, we asked 31 participants (who did not take part in the main experiment or the first behavioural study) to provide a sentence continuation after hearing each unique SNP+verb in our stimuli (e.g. "*The experienced walker chose* …"), for example "… the shorter route", "… the hardest path" etc. Note that we only used the noun responses which are considered to be an object of the preceding verb (e.g. nouns in DO or PP complements which we refer to as CNs throughout this paper) in order to remove any syntactic or thematic variability when modelling semantic interpretation of the CN. For example, any noun responses in a sentential complement were removed since they are often treated as a new subject instead of an object (e.g. "*The walking couple heard that the* ***farm*** *was open to visitors*"). On average, this left 18 CN responses for every stimulus from 31 participants. Any stimulus with less than 4 responses were excluded from the analysis.

3.2) Semantic modelling

We trained a probabilistic topic model based on LDA (Griffiths and Steyvers 2004). It develops a generative probabilistic model which assigns a word to different latent dimensions in a way that maximizes the posterior of the model. Such latent dimensions are often called "topics" which describe the semantic content of a word in the form of a probability distribution. In this study, topic distributions (consisting of 100 topics) associated with each content word were generated using corpus-based tensor data (Baroni and Lenci 2010). Instead of using raw co-occurrence frequency, we used local mutual information (LMI) from the tensor because it normalises the effect of lexical frequency of individual items when computing the semantic relation (co-occurrence) between two words. Further, instead of using all co-occurrence data in the tensor, we only selected specific subsets in order to capture syntactically licensed semantic representation specifically with respect to a word in the context. In particular, we focused on the incremental and cumulative development of the semantic constraint from a SN (agent) to a CN. To do this, we trained two separate topic models based on the co-occurrence between 1) SN and verb (SN-V) and 2) the preceding words including SN and verb and CN (object) (SNV-CN). These models provided different aspects of semantic

representation relevant for incremental predictive processing as follows: 1) the first (SN-V) topic model was trained specifically to characterize the predictive representation of SNs on upcoming verbs and the specific semantic content of verbs that are syntactically licensed with respect to the preceding SNs and 2) the second (SNV-CN) topic model was trained specifically to characterize predictive representation of SNs and verbs on CNs and, the specific semantic content of CNs that is syntactically licensed with respect to the preceding SNs and verbs. See Section 1 in Supplementary Information for more details regarding model training and parameter settings.

3.3) Modelling predictive state: semantic blends

After obtaining the behavioural responses from the two sentence completion studies (verbs from the first and CNs from the second study) and topic representation associated with a set of unique responses for each sentence, we combined them to generate an overall representation across multiple responses (for either the unique verbs or the CNs) to capture consistent semantic content shared by the set of verbs predicted by the SNP or by the CNs predicted by the SNP+verb. In this way, we aimed to model predictive activation of semantic contents associated with multiple lexical items based on the preceding context. The semantic blend was computed as below:

$$blend(words) = P(topic \mid full\ context) = \sum_{word} P(topic \mid word)P(word \mid full\ context)$$

where $P(word \mid full\ context)$ is a probabilistic weight associated with a given $word$ (see 3.1) and $P(topic \mid word)$ is the topic distribution for $word$ (see 3.2). Based on this formula, we constructed three different "blend" vectors.

a) SN-V verb blend

This blend is designed to model the SNP constraint on upcoming verbs. We counted the (post-SNP) verb responses from the first sentence completion study. Then the frequency count associated with each unique verb that was produced by participants was, in turn, used as a weight to the topic distribution of the verb. From the topic model trained specifically on the SN-verb co-occurrence data, we obtained the topic representation of each unique verb which was weight-combined as expressed in the formula above (i.e. $P(verb\_topic \mid verb)P(verb \mid SNP)$).

b) SNV-CN verb blend

Despite being a verb blend, this second blend model is designed to model the SNP constraint on CNs (rather than its constraints on the verb), via the set of predicted verbs obtained from the first behavioural study. We counted the (post-SNP) verb responses and the frequency count associated

with each unique verb that participants produced as above. However, we obtained the verb topic distributions from a second topic model trained specifically on the mixed SN-CN and verb-CN co-occurrence data, reflecting the predictive representation on upcoming CNs. Then, each predictive representation (topic-context distribution) of unique verbs in relation to CNs was weight-combined as expressed in the formula above (i.e. $P(CN\_topic \mid verb)P(verb \mid SNP)$)

c) SNV-CN CN blend

The third blend focused on modelling the combined constraint of SNP+verb on CNs. To do this, we counted the (post-SNP+verb) CN responses from the second sentence completion study. Then, we used the CN topic distributions from the second topic model trained specifically on the mixed SN-CN and verb-CN co-occurrence data, reflecting the topic representation of each unique CN in relation to the preceding subjects and verbs. Then, just as the other blends, each topic representation (target-topic distribution) associated with each unique CN was weight-combined as expressed in the formula above (i.e. $P(CN\_topic \mid CN)P(CN \mid SNP + verb)$).

In summary, we generated the following blends whose entropy is designed to address how constraints incrementally change and develop:

a) $P(verb\_topic|SNP) = \sum_{verb} P(verb\_topic|verb)P(verb|SNP)$
b) $P(CN\_topic|SNP) = \sum_{verb} P(CN\_topic|verb)P(verb|SNP)$
c) $P(CN\_topic|SNP + verb) = \sum_{CN} P(CN\_topic|CN)P(CN|SNP + verb)$

3.4) Modelling predictive constraint: entropy

Entropy is a metric designed to quantify the amount of uncertainty in distributional models. Therefore, entropy of the blend distributions in this study reflects the strength of semantic constraint regarding upcoming words (higher uncertainty=weaker constraint). However, in any topic models, each topic varies in terms of the types of words it prefers with different probabilities. This naturally leads to variations in semantic dispersion across topics, potentially undermining the estimation of true semantic entropy. Here, we addressed this issue by linearly combining entropy with topic dispersion as following:

$$H(P(x)) = w * \mathrm{h}(P(x)) = \sum_i w_i[\, -P(x_i)\log P(x_i)\,]$$

where w is a vector of semantic dispersion across topics and $\mathrm{h}(P(x))$ is a vector containing local entropy values. In this paper, we denote the term entropy and notation H to refer to this dispersion-corrected entropy. The semantic dispersion was calculated by averaging pair-wise cosine distances

between topic distributions among every pair of words within a topic (Lyu et al. 2019). If the target words preferred by a topic have similar distributions, the average cosine distance will be low. Then, this "within-topic" semantic dispersion was linearly combined with the local entropy values to manipulate the contribution of each topic to the degree of overall constraint strength across topic candidates. In this way, we effectively controlled for "within-topic" dispersion when computing "between-topic" constraint.

<INSERT TABLE1>

Each of the semantic blends described above was taken as an input to the entropy function (Figure 3), generating three semantic constraint models which were tested against the spatiotemporal patterns of neural activity at specific epochs (Table 1):

1) SNP adjacent constraint model on upcoming verbs: entropy of $P(verb\_topic|SNP)$

2) SNP non-adjacent constraint model on upcoming CNs: entropy of $P(CN\_topic|SNP)$

3) SNP+verb constraint model on upcoming CNs: entropy of $P(CN\_topic|SNP + verb)$

3.5) Modelling evaluation: constraint mismatch

Semantic evaluation refers to a process of resolving mismatch between a current input and the predicted candidates based on the preceding context, leading to an accurate interpretation of the input that fits the context. To model this process, we quantified the degree of mismatch by computing cosine distance between the semantic representations of the predicted CNs and the target CN. As described in 3.1, we excluded any items that do not contain CN (i.e. a noun considered to be an object of a preceding verb) from the analysis because this mismatch model requires the target CN to be identified. This left us with 128 out of 200 trials.

4) ***Spatiotemporal searchlight representational similarity analysis***

In order to determine when and where these constraint models and associated computations are neurally realised, we used spatiotemporal searchlight Representational Similarity Analysis (ssRSA) (Su et al. 2012). Each searchlight is defined for each vertex at each time-point, providing a fine-grained spatiotemporal map of neural activity. To characterise such dynamic pattern of neural activity, we constructed model representational dissimilarity matrices (RDMs) using specific properties of the blended distributions across sentences described above. Since all of the model RDMs in this study were based on the summary metrics designed to capture various incremental aspects of distributional semantics, the representational geometry was characterized simply by

calculating absolute distance of the metric values between every pair of trials. Each of these model RDMs was, then, compared with the patterns expressed by the neural RDMs constructed by correlation distance between every pair of trials for each searchlight across space and time (see Figure 4). The size of each searchlight was set as a spatial radius of 10mm and temporal radius of 30ms.

ssRSA was performed within a language mask, which included all anatomical regions in a set of regions encompassing bilateral fronto-temporo-parietal regions, using the Harvard-Oxford cortical atlas (Kocagoncu et al. 2017; Lyu et al. 2019). See Figure 4 for surface rendering of this language mask. These regions are reliably shown to be involved in language processing (Binder et al. 2009; Price 2010, 2012).

5) EMEG recordings and MRI acquisition

MEG data were recorded on a VectorView system (Elekta Neuromag, Helsinki, Finland) using 306 sensors (102 magnetometers and 204 planar gradiometers), located in a magnetically shielded room at the MRC Cognition and Brain Science Unit, Cambridge, UK. In conjunction with the MEG recordings, we recorded EEG signals using an MEG compatible EEG cap (Easycap, Falk Minow Services, Herrching-Breitbrunn, Germany) with 70 electrodes, plus external electrodes and a nose reference. To monitor head movement in the MEG helmet, five HPI (Head Positioning Indicator) coils attached to the scalp recorded head position every 200ms. Blinks and eye movements were recorded by EOG (Electro-Oculogram) placed above and beneath the left eye and beside the left and right outer canthi. Cardio-vascular effects were recorded by ECG (Electro-Cardiogram) attached to right shoulder blade and left torso. To be able to co-register the EEG and MEG data to anatomical structural scans for each participant, the positions of the HPI coils and EEG electrodes were digitised relative to three anatomical landmarks (nasion, left and right peri-auricular points). In addition, a participant's head shape was digitised across the head. MEG signals were recorded with a sampling rate of 1000Hz and any signals below 0.03Hz were high-pass filtered.

To localise the EEG and MEG data to sources on the cortical surface, structural MRI scans were acquired for each participant in a separate session using 1mm isotropic resolution T1-weighted MPRAGE on a Siemens 3T Prisma scanner (Siemens Medical Solutions, Camberley, UK) located at the Cognition and Brain Science Unit, Cambridge, UK.

6) EMEG pre-processing

The raw MEG data were max-filtered (Elekta-Neuromag) to remove bad channels, to compensate for head movement using Signal Space Separation (SSS) techniques (Taulu and Simola 2006).

SPM8 (Statistical Parametric Mapping 8, Welcome Institute of Imaging Neuroscience, London, UK) was used to complete the remaining stages of EMEG preprocessing (except for independent component analysis (ICA) artefact rejection). First, a low-pass filter at 40Hz was applied to the data using a 5^{th} order bidirectional Butterworth Digital Filter. In order to remove any physiologically driven artefacts such as blinks or cardiac signals recorded by EOG and ECG, the data signals were decomposed into independent components (IC) and each IC was correlated with vEOG, hEOG and ECG channels. Any ICs showing very high temporal correlation (correlation > 0.3) with any of these channels were removed and the remaining ICs were then visually inspected to ensure that no artefact component remained. The remaining ICs were then used to reconstruct the data.

Next, three separate analysis epochs were generated by aligning the data to the onset of each of the three points of interest in each sentence (see Figure 1). The duration of each epoch (0 to 600ms) was consistent across all three epochs. This duration was chosen to cover the average duration of each word + one standard deviation described in Figure 1. One epoch was aligned to the SN, another to the verb and a third to the CN. We also calculated the uniqueness point (UP) of each of these words from CELEX database (Baayen et al. 1993) to relate the timing of neural effects to when the word is recognised.

After epoching, each channel was baseline-corrected by subtracting the time-averaged data from a baseline period -200ms to 0ms relative to sentence onset (i.e. a period of silence immediately preceding the sentence). Finally, automatic artefact rejection was used to identify trials for which 15% or more sensors in any one of the three sensor types exceeded amplitude threshold (6e-11T for magnetometers, 3e-12T/m for gradiometers and 2e-04V for EEG), and these trials were rejected (an average of 15 trials were rejected (SD = 13.43)).

7) EMEG source reconstruction

Source reconstruction aims to estimate the regional response within a brain using the EMEG data recorded outside the scalp. We first transformed the participants' structural MRI images into an MNI template brain which was then inverse-transformed to construct individual scalp and cortical meshes by warping canonical meshes of the MNI template brain to the original MRI space (Mattout et al. 2007). The MRI co-ordinates from individual scalp and cortical meshes were co-registered with the MEG sensor and EEG electrode co-ordinates by aligning fiducial points and the digitised head shape to the outer scalp mesh. A single-shell conductor model and a boundary element model (BEM)

were used as forward models for MEG and EEG recordings respectively (the defaults in SPM8). We source-reconstructed our data based on the minimum-norm assumption in SPM8 as a prior on the source covariance (López et al. 2014). This source prior was empirically adapted to maximise the model evidence which, in turn, was used to compute the Maximum A Posterior (MAP) source estimate.

8) Statistics and multiple comparisons correction

Using the correlation time-courses for the model and data RDMs across subjects, we calculated a time-course of one-tailed t-statistic for every vertex (Figure 4). From this point-wise statistic, we applied the cluster forming threshold (CFT) of p=0.01 and binarized the time-courses into clusters from a set of temporally and spatially contiguous vertices (data-points). Then, we summed t-values across each of the vertices within a cluster to compute a cluster-summed t-value. In this way, we aimed to emphasize the neural clusters that are spatiotemporally distributed while each of the vertices in the clusters shows p-value less than 0.01.

For multiple comparisons correction across time-points which are not independent of one another, we ran permutation statistics (Maris and Oostenveld 2007) on the CFT output. Under the null hypothesis that our model is not correlated with the data (r=0), we randomly permuted the sign of correlation values across different subjects and ran one-sample t-test for every time-point. For each randomization, this null time-course of t-values was converted to the time course of cluster-summed t-statistics. This random permutation process was repeated 1,000 times and the cluster with the maximum t-value across all data-points for every run was saved. This process gives 1,000 cluster-level t-values under the null hypothesis and the significance of the observed cluster-level t-values were evaluated with respect to this null distribution.

RESULTS

Using RSA and model RDMs of semantic constraint and mismatch, we probed source-localised EMEG data capturing the real-time electrophysiological activity of the brain to determine the spatiotemporal properties of the cumulative incremental effects of semantic constraints. For this purpose, we directly compared the strength of semantic constraints generated by the SNP on verbs and CNs, as quantified by the entropy of $P(verb\_topic|SNP)$ and $P(CN\_topic|SNP)$, against the multivariate patterns of neural activity over space and time. Then, we looked at the effects of the combined SNP+verb constraint by computing entropy of $P(CN\_topic|SNP + verb)$. In this way, we

aimed to investigate the timing and neural regions that are related to generating semantic constraints prior to a target word (i.e. verb or CN). Lastly, to measure the predictive effects of the incrementally developed constraint on the processing of the CN semantics, we constructed a constraint mismatch model to examine the neural effects of semantic evaluation. We report significant ($p \leq 0.05$) and marginally significant ($0.05 < p \leq 0.06$) effects of the models sequentially as the sentence unfolds over time. Note that all of these reported results have large effect sizes ($d > 0.8$; see Supplementary Information Section 2).

(a) SNP's adjacent semantic constraint (entropy) on upcoming verb

We anticipated that the semantics of the SNP (e.g. "*The experienced walker*") would generate rich constraints on the upcoming speech. To test this hypothesis we constructed models capturing the strength of constraints generated by the SNP (e.g. entropy of $P(verb\_topic|SNP)$ in this section and $P(CN\_topic|SNP)$ in the section below). Using these entropy models, we aimed to assess the earliness of predictive computations and how they develop throughout a sentence. The results (Figure 5a) show that the constraints on the verb generated by the SNP are significantly activated around the UP (347±107ms after the onset) of the SN as it is recognised, lasting around 300ms from 290 to 600ms and are seen primarily in right hemisphere (RH) mid-anterior middle and inferior temporal areas ($p = .032$). This effect continued until the end of the SN (Epoch 1) and was not significant in Epoch 2, suggesting that listeners are actively constraining upcoming verbs as soon as they recognise the SNP and that these constraints involve only RH temporal regions.

(b) SNP's non-adjacent semantic constraint (entropy) on CN

When examining constraints on non-adjacent words in a sentence (in this case SNP constraints on the CN), we need to consider the semantic relation between the context (SNP) and the target (CN) while taking into account any words that intervene between them (in this case, the verb). Using the Bayesian approach, we computed the non-adjacent SNP constraint on CNs by taking into account the set of verbs that were predicted by hearing the SNP in the first behavioural completion study: $\sum_{verb} P(CN\_topic \mid verb) P(verb|SNP)$. This mathematical formulation reflects the SNP constraint on CN semantics via the set of verbs predicted by the SNP collected from the first behavioural study. This set of predicted verbs can be thought of as a process of semantic competition amongst partially activated semantic candidates. This is similar conceptually to the notion of cohort competition for spoken language comprehension (see (Marslen-Wilson 1987) which claims that multiple, partially

activated word candidates initiated by the accumulating speech input as a word is heard momentarily compete with each other until the word is recognised). Applying topic modelling to these predicted verbs enables us to model the SNP's constraints on the CN taking into account the scope of the SNP's prediction on the intervening verb.

Similar to the SNP's constraint on verbs, this non-adjacent constraint appeared around the UP of the SN starting from 270ms to 590ms after the SN onset (Figure 5b). It involved early, relatively short-lived effects in bilateral anterior and middle temporal cortex (left hemisphere (LH): p=.026 from 280ms to 510ms; RH: p=.039 from 280ms to 530ms), which overlapped with effects in right inferior frontal areas (p=.026 from 270ms to 590ms; see Figure 5b). Note that these are the results from Epoch 1 aligned to the SN onset.

In a further analysis we tested the spatiotemporal patterns of neural activity with the same non-adjacent SNP constraint model in Epoch 2 (Figure 5b). We found a significant SNP semantic constraint effect on the CN but only in the right infer frontal gyrus (RIFG) from the verb onset (p=.01; Figure 5), lasting for 380ms (one standard deviation after the mean UP), suggestive of competitive processing. We discuss the differential role of RIFG from the RH temporal regions in light of the constraints that they activate in the Discussion.

(c) SNP+verb's semantic constraint (entropy) on CN

The analysis above examined the effect of the constraints imposed by the SNP on the CN mediated through verbs predicted in the behavioural test. In this section we investigate the changes in the semantic constraint on CN as the SNP context becomes enriched by combining with its adjacent verb (i.e. after the cohort competition among the verb candidates has ceased, a process reflected in the blend model). To do this, we tested the effect of the SNP+verb constraint model on CNs (i.e. entropy of $P(CN\_topic|SNP + verb)$), in order to elucidate the neurobiological basis of the development of incremental constraints (Figure 5c). Our results showed that right mid-anterior middle and inferior temporal areas again played a role in constraining the CNs from 60ms after the verb onset and lasting around 500ms (p=.002; Figure 5c). This early constraint effect likely reflects the constraint driven by the event generated by the SNP, which could be largely consistent with the constraint imposed by the verb, especially when the verb is light in terms of its semantic constraint as in the majority of our sentence stimuli (see Discussion). In addition, we also found a significant cluster in left anterior middle and inferior temporal regions from 270ms to 470ms (p=.025) and a marginally significant cluster in left inferior frontal gyrus (LIFG (BA47/45); p=.06). Based on the involvement of

LATL and LBA47/45 in constraining upcoming CNs around the UP of the verb, we speculate that their role is to unify the verb into the broad semantic constraint set up by the SNP, essentially leading to a reduction in uncertainty in the constraint (see Figure 3 and Section 3 in Supplementary Information). The effect of the combined SNP+verb constraint persisted into epoch 3 during which the CN is heard. This lasted until 370ms into the CN which is around the UP. However, this transition was associated with more posterior LH regions in middle temporal gyrus (MTG) and angular gyrus (AG) (p=.031; Figure 5c). This anterior-to-posterior transition may underscore the process from constructing to utilizing the context-driven semantic constraint when hearing the CN in a sentence.

(d) Semantic mismatch between the target CN and the predicted CNs by the SNP+verb context

Our final analysis was aimed at demonstrating how the prior SNP+verb constraint facilitates the interpretation of the CN in light of its preceding context. To do this, we computed the cosine distance between topic representation of the target CN (i.e. $P(CN\_topic|CN)$) and blend representation across the predicted CNs by the preceding SNP+verb context (i.e. $P(CN\_topic|SNP + verb)$). This model reflects the degree of mismatch between the predicted and the target semantics of the complement. This measure can be viewed as an index of semantic evaluation as it indicates the difficulty of processing the CN in light of the preceding context. Using this model, we observed a cluster (marginally significant (p=.058) in LH posterior middle temporal gyrus from 370ms to 520ms after the CN onset (Figure 5d). The timing of this mismatch effect emerges just after the constraint effect disappears, suggesting that the constraint is evaluated against the CN as soon as the predictive process terminates and the CN is fully identified. This last piece of evidence sheds light on the predictive computations actively engaged by listeners while incrementally processing the subject, verb and object, which are critical components of understanding the message that speaker conveys.

Discussion

The goal of the present study was to understand the neural dynamics of cognitive processes as listeners incrementally interpret the spoken sentences that they hear. The computations involved in this process include (1) the activation of the semantic constraints generated by the semantic content of each word in a sentence as it is heard based on activated broad scenarios [or event structures], (2) how and when these constraints affect processing of the upcoming speech, and (3) the incremental fine-tuning and evaluation of the semantic constraint on each new word, integrating it into the developing semantic representation. During the experiment, listeners heard sentences consisting of

a SNP, followed by a verb, and then a CN where the SNP and the verb varied in the cumulative probabilistic constraints they generated on the upcoming complement. We tested for the timing and neural location of these computations by recording real-time brain activity using EMEG and analysing the spatiotemporal fit of patterns of probabilistic topic models with source-localised neural activity across an extensive set of bilateral frontal, parietal, and temporal regions.

Our summary of the results with respect to the timing of effects throughout the entire sentence reveals the rapid transitions of information processing in the brain as each word (SN, verb, and CN) incrementally unfolds over time. Such transitions highlight the underlying neural computations not only involved in processing individual words, but also in combining them with the prior context to develop a representation of the meaning of the sentence. More specifically, our results revealed the spatiotemporal dynamics of incremental semantic computations in the brain: 1) The early activation of semantic constraints generated by the SNP primarily engaged RH mid-anterior temporal areas whereas activating the non-adjacent constraint on CNs additionally recruited the RIFG and left temporal regions; 2) As the verb is recognised, the RH clusters started to decline but new clusters emerged in anterior left IFG (LIFG) and left anterior temporal lobe (LATL), actively constraining CNs based on the combined SNP+verb context; 3) As the target word (CN) starts to be heard, the locus of the SNP+verb constraint moved posteriorly into the left posterior MTG (LpMTG) and LAG which lasted until the CN is recognised. Here, we discuss our results in relation to incremental processing issues from the SNP to the CN (see Figure 6).

Early activation of the SNP constraints

Our results revealed that different aspects of SNP constraints are activated between the point at which the SNP is recognised (i.e., the UP of SN) and its offset approximately 100ms later and that these computations recruit different brain areas. First, the SNP constraint on upcoming verbs (Figure 5a) appeared only in mid-anterior portions of right middle/inferior temporal gyri (RMTG/ITG) whereas the SNP constraint on upcoming CNs (Figure 5b) involved more extensive regions including right ATL (RATL), RIFG and LH temporal cortex. The important similarities and differences in the neurobiological basis of these constraints are (1) the core regions involved in constructing both types of constraints which included RH anterior MTG/ITG regions and (2) only the non-adjacent SNP constraint on CNs elicited activation in the RIFG which lasted all the way until the verb was recognised in Epoch 2.

These regions are plausibly involved in generating and maintaining the event representations which are naturally generated at the beginning of sentences and form a basis for semantic constraints on upcoming speech (Marslen-Wilson et al. 1993; Nieuwland and Van Berkum 2006). Various studies (Marslen-Wilson and Tyler 1980; Kamide et al. 2003) have shown that listeners use multiple sources of information at the earliest possible opportunity to establish the fullest possible interpretation of what they are hearing and demonstrates such processes are not restricted to the syntactic structure of language. One of the prediction principles (Altmann and Mirković 2009) that underpin human language comprehension states that the mapping between the unfolding sentence and the event representation enables listeners to predict both how the language will unfold and how the real-world event will unfold, rendering prediction impossible to stand alone without incrementally developing event representations.

In line with these claims, our results revealed consistent activations of RH mid-anterior temporal regions for different semantic constraints, likely reflecting the broad scenarios activated by the SNP. This claim is further supported by three major findings from our main and complementary analyses, possibly indicating that they are activated from the same set of scenarios drawn by the SNP: (a) the same activation timing for different SNP constraints around the UP of a SN, (b) a common subspace existing between different SNP constraints (see Section 4 in Supplementary Information) and c) the joint semantic constraint of the SNP on verb and CN (i.e. the early event-level constraint) elicited a significant activity pattern in the RH mid-anterior temporal regions as well (see Section 5 in Supplementary Information).

The activation of RH regions have been consistently reported when drawing coherent “message-level” interpretations in speech comprehension (Beeman and Chiarello 1998; Beeman et al. 2000; Jung-Beeman 2005), consistent with studies claiming the importance of RH in processing linguistic context (Kircher et al. 2001; Bookheimer 2002). These findings have been supported by previous ERP studies showing that the RH plays an important role in interpreting individual words with respect to a larger-scale context (Federmeier and Kutas 1999; Wlotko and Federmeier 2007; Federmeier et al. 2008), emphasising the role of RH in processing context-driven semantic relationships (Federmeier et al. 2008). Hence, the early effect in the right temporal regions in the current study are likely related to the process of generating constraint driven by the SNP context, setting up the event-level scenarios of what is likely to be talked about (Elman 2011).

However, two additional areas in the LH temporal lobe and RIFG were engaged in constraining the non-adjacent CN based on the SNP context (Figure 5b). The two critical differences between the SNP constraints are 1) the grammatical category of constrained words and 2) adjacency with respect to

the SNP context. Previous studies have shown the engagement of the LH temporal regions when processing nouns compared to when processing verbs (Siri et al. 2007; Vigliocco et al. 2011). Unlike the bilateral temporal regions, the RIFG cluster remained significant after the verb onset until the verb was recognised. Consistent with this finding, recent studies have reported RIFG as a part of the extensive network involved in constraining an upcoming word (Willems et al. 2015) and resolving semantic competition (Kocagoncu et al. 2017). More generally, this region has been involved in semantic maintenance and cognitive control (Shivde and Thompson-Schill 2004; Gajardo-Vidal et al. 2018), activating when processing an indeterminate sentence which can be interpreted in many different ways (de Almeida et al. 2016) or when encountering a word with multiple meanings in a spoken sentence (Rodd et al. 2005; Mason and Just 2007). Therefore, the SNP constraint effect in RIFG during the verb likely reflects maintenance of the SNP semantic constraint while resolving competition as the verb is being heard.

Evolving constraint

The essence of incremental speech comprehension is that each word is interpreted in a context-relevant manner and the constraint derived from the prior context is updated to be more specific and informative on the upcoming words in the sentence as more words are heard (Kuperberg and Jaeger 2016). To investigate this incremental development (i.e. how the prior SNP constraint on CNs evolves as a verb is recognised) we constructed a model that captures the semantic constraint on CNs based on the full SNP+verb context. Our results showed that the effect of the SNP+verb constraint appears at 60ms after the verb onset in the right mid-anterior MTG/ITG regions which extended to LATL and LIFG peaking around 400ms after the verb onset (i.e., close to the mean verb offset). As the target word (i.e., CN) is being heard, the cluster moved into more posterior areas involving LMTG and LAG which lasted until the CN is recognised (Figure 5c). These transitions across time may highlight differential roles engaged by these regions when constraining the CN. For example, as discussed above, the early RH temporal effect most likely reflects the broad constraint on CN, primarily set up by the SNP (i.e. in natural language comprehension, it is highly unlikely that an incoming verb is completely incongruent with the activated scenarios). Then, the ventral fronto-temporal network in LH including LIFG (BA47/45) and LATL additionally engages in constraining the CN as the verb is recognised.

The broad scenarios activated by the SNP become more fine-tuned as the semantics of the verb is combined with the SNP context. According to the timing of LIFG-LATL activations, these regions may play an important role in resolving uncertainty by updating the sentential meaning so that it becomes more specific. Further support for this argument comes from a complementary analysis

(see Section 3 in Supplementary Information) showing a statistically significant reduction in entropy between the SNP constraint and the SNP+verb constraint, which reflects an important aspect of incremental speech comprehension (Hale 2006) (see Figure 3). As LATL is directly connected to LBA47 via the uncinate fasciculus (Catani et al. 2005), our results suggest that the interaction within the anteroventral fronto-temporal network is involved in developing more informative constraint based on the combined context of SNP+verb.

After the onset of the target word (CN), we observed a significant cluster moving into more posterior regions including LpMTG and LAG until around the UP of the CN. The transition and timing of this cluster may reflect the facilitatory effect of the contextual (SNP+verb) constraint on activating semantic content of the CN as these regions are often involved in activating lexical-semantic content (Hickok and Poeppel 2007) and combining it into the preceding context at both phrasal and sentential levels (Humphries et al. 2007; Schell et al. 2017; Lyu et al. 2019). Therefore, such anterior (BA47/45 and LATL) to posterior (LpMTG/LAG) transition likely reflects the top-down (i.e. the SNP+verb constraint) bottom-up (i.e. speech input of the CN) interaction, in order to generate a coherent semantic interpretation of the CN with respect to the preceding SNP+verb context.

Constraint evaluation

Developing an event representation requires each word in a sentence to be interpreted in the context of the prior context. This process, in turn, requires semantically evaluating each word with respect to the prior constraint, indexed by the degree of mismatch between the context and an upcoming word. To address this issue, we tested the effect of contextual (SNP+verb) constraint on the interpretation of the target word (CN) by quantifying the degree of mismatch between the sentential context and the target word in terms of the spatiotemporal patterns of neural activity after the CN onset. We found that activity patterns in LpMTG were sensitive to the mismatch between the constrained and the actual topic representation from 370ms to 520ms (Figure 5d). Interestingly, this timing occurred immediately after the constraint effect disappeared.

In the literature, LpMTG is commonly reported in studies of semantics (Price 2010) and is typically known as the source of the N400 effect (Lau et al. 2008; Kutas and Federmeier 2011). A recent study reported predictability (e.g. “runny nose” vs. “dainty nose”) estimated from corpus data modulated the N400 component in LpMTG (Lau and Namyst 2019), reducing the necessity of activating the stored lexical representation of the target word (CN in our study) when it is strongly constrained by the context (i.e. high predictability).

This argument is further supported by our previous study (Lyu et al. 2019) where the semantic representation of a CN was strongly modulated by the preceding verb; for example, the verb in

context (e.g. the man “ate”) pruned the less relevant CN topics, allowing listeners to interpret the CN (e.g. “apple”) more specifically with the CN topics that were supported by the preceding verb (e.g. topics related to “food” but not those related to “shape” or “colour”). While the exact computational details of the mismatch effect remain elusive, our findings suggest that listeners not only develop semantic constraints on upcoming words but they also use these constraints to efficiently derive the context-relevant interpretation of upcoming words like the CN. Combined with other constraint effects discussed above, these results clearly illustrate the incremental stages of predictive processing that enables listeners to construct the message-level interpretation from the three crucial components in a sentence (SNP, verb and CN).

Implications for future studies

Previous studies have explained neuroimaging data using computational models to quantify entropy at lexical (Frank et al. 2015; Willems et al. 2015) and phonological levels (Donhauser and Baillet 2020). In these studies, neural network models with a recurrent architecture were commonly employed to generate a context-dependent linguistic prediction as a probability distribution from which entropy can be computed. On top of these studies, the current study examined the semantic aspect of incremental language prediction using entropy of topic distributions, designed to express the co-occurrence relation among words in different grammatical categories through estimating the expected posterior of the multinomial parameters (see Supplementary Information section 1). In this section, we motivate the choice of our computational model and approach while discussing its limitations and directions for future studies.

Recent advances in the field of computer science have established a number of different computational algorithms to construct distributional semantic models (DSMs), optimally reflecting the content of each lexical item in a set of latent dimensions. Perhaps, the currently most popular algorithm is the neural network training with a recurrent architecture including recurrent neural network (RNN) and long short-term memory (LSTM). However, we chose to use topic modelling based on LDA to exploit its two critical aspects:

1. It produces a semantic vector of a word as a probability distribution over latent semantic dimensions (topics). This allows us to construct our incremental models under the Bayesian computational framework (Kuperberg and Jaeger 2016), a useful approach for understanding predictive processing in language.
2. It explicitly depicts the semantic relations between words in different positions in a sentence. Our implementation of topic modelling, which treats SNs and verbs as

"documents" and CNs as "words", is specifically designed to explain semantic prediction and updates based on key words in the context.

Its explanatory value as a predictive model is one of its biggest assets, making it particularly attractive in the field of psycho- and neuro-linguistics. Nonetheless, one critical limitation of this approach is that it is not an incremental model by itself, unlike RNN or LSTM. To address this issue, we introduced the method of blending a set of topic vectors based on Cloze probabilities calculated from sentence completion studies.

Despite the popularity of Cloze probability as a direct behavioural measure of human prediction, its application entails high subjective bias, often affected by confounding factors such as familiarity (Smith and Levy 2011). Although Cloze probability was significantly related to corpus probability, it also significantly deviated from the corpus probability with greater entropy in responses, making Cloze a suboptimal estimate of linguistic prediction which has been successful in explaining neural responses (DeLong et al. 2005; Kutas and Federmeier 2011). Moreover, another confounding factor of Cloze is that the prediction may well be driven by a pragmatic inferential process, not purely by semantic associations. Hence, it remains controversial whether the basis of the incremental prediction is semantic or pragmatic in nature. Despite the objective and accurate probability estimates that large-scale corpora offer, there is a practical limitation of applying the corpus probability as the number of words increases in the model (i.e. increasing N in an N-gram probability). Even with large-scale corpora, the estimation of co-occurrence probability becomes very difficult with $N > 3$. With our stimuli containing 6-7 words before the complement noun (e.g. "The experienced walker chose the path"), computing a conditional probability becomes impossible. Taken together, future studies need to develop a self-explanatory incremental model, allowing us to characterize evolving representations. Recent developments of more sophisticated models such as generative pre-training (Radford et al. 2018) have shown impressive performance on making output predictions but their multi-layered internal representations are highly complex and lack an explanatory value to provide insights into predictive processing in the human brain. Quantifying different aspects of representation that incrementally evolve over time in these models will initiate more model-driven decoding research on brain data, shedding light on the neurobiological basis of incremental speech comprehension. Lastly, although we constrained our search space within a language mask to characterize linguistic aspects of predictive processing and specific computations involved in constraining upcoming words, other brain networks involved in different cognitive functions such as attention and/or memory may also be involved in such linguistic processes of understanding speech. With the ultimate goal of expanding our research to discourse and narrative

comprehension, such whole-brain analysis will contribute to understanding the interactive nature of cognitive processes during language comprehension.

Finally, there have been growing efforts to elucidate the interactive nature of cognition, bringing multiple domains of cognition such as language and memory into a unifying framework (Duff and Brown-Schmidt 2017). For example, developing an event representation involves the episodic realization (e.g. “orange” in “She peeled an orange, and ate it quickly”) of a semantic type (e.g. “orange” in general). The role of such episodic-semantic interface during natural language comprehension is extensively discussed in a recent account (Altmann 2017), claiming the hippocampal structures as one of the neurobiological bases for encoding distinct episodes (McClelland et al. 1995). While we have shown that incremental predictive processes can be characterized even with such generic linguistic stimuli, we advocate the need for more specific stimuli in a narrative context in order to distinguish an episodic token from a semantic type. In this way, the stimuli would have sufficient variability to provide the distinguishable representational geometry between them, allowing researchers to investigate the interactive event dynamics beyond combinatorial semantics from semantic memory alone.

As a final remark, this study focused on presenting a possible approach to investigate one of the core processes (i.e. prediction) of human event cognition during natural speech comprehension. Future studies will need to expand this research to investigate other central cognitive processes involved in understanding the event dynamics and illuminate its neurobiological underpinnings, likely recruiting multiple interactive networks in the brain outside the language network.

Conclusion

In this study, we demonstrated the neurobiological basis of incremental predictive language processing by characterising the spatiotemporal dynamics of source-localised EMEG data with ssRSA using rich co-occurrence computational semantic models based on topic modelling combined with human behavioural data.

To summarise our results, an extensive bilateral fronto-temporo-parietal network is actively engaged in generating and developing incremental semantic constraints on upcoming words (see Figure 6). Our results highlight the temporal progression of semantic constraint development: (1) A RH fronto-temporal network initially generates possible scenarios as the SNP is heard which, in turn, (2) recruits a LH fronto-temporal network as the scenarios get enriched as subsequent words are heard (a verb in this case) and, (3) terminating in a LH posterior temporo-parietal network as the target word (CN) is recognised. To our knowledge, none of the neurobiological models of speech comprehension have explained this range of sequential temporal relationships among multiple

regions in the language network during incremental speech comprehension, largely due to the lack of evidence for characterizing the spatiotemporal dynamics of neural activity. Further research is needed to understand the detailed neural mechanisms underpinning these important effects.

Correspondence should be addressed to Prof. Lorraine K. Tyler, Centre for Speech, Language and the Brain, Department of Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK. Tel: 01223 766457, E-mail: lktyler@csl.psychol.cam.ac.uk.

Acknowledgements

This work was supported by a European Research Council Advanced Investigator Grant to L.K.T. under the European Community's Horizon 2020 Research and Innovation Programme (2014-2020 ERC Grant Agreement 669820). We thank Dr Barry Devereux for his help in the early stages of this research.

References

Altmann GTM. 2017. Abstraction and generalization in statistical learning: implications for the relationship between semantic types and episodic tokens. Philos Trans R Soc B Biol Sci. 372:20160060.

Altmann GTM, Mirković J. 2009. Incrementality and prediction in human sentence processing. Cogn Sci. 33:583–609.

Baayen RH, Piepenbrock R, Van Rijn H. 1993. The CELEX lexical database (CD-ROM). Linguistic data consortium. Philadelphia, PA Univ Pennsylvania.

Baroni M, Lenci A. 2010. Distributional memory: A general framework for corpus-based semantics. Comput Linguist. 36:673–721.

Beeman MJ, Bowden EM, Gernsbacher MA. 2000. Right and left hemisphere cooperation for drawing predictive and coherence inferences during normal story comprehension. Brain Lang. 71:310–336.

Beeman MJ, Chiarello C. 1998. Complementary right-and left-hemisphere language comprehension. Curr Dir Psychol Sci. 7:2–8.

Bicknell K, Elman JL, Hare M, McRae K, Kutas M. 2010. Effects of event knowledge in processing verbal arguments. J Mem Lang. 63:489–505.

Binder JR, Desai RH, Graves WW, Conant LL. 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb Cortex. 19:2767–2796.

Bookheimer S. 2002. Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. Annu Rev Neurosci. 25:151–188.

Bornkessel-Schlesewsky I, Schlesewsky M. 2013. Reconciling time, space and function: a new dorsal–ventral stream model of sentence comprehension. Brain Lang. 125:60–76.

Campbell KL, Tyler LK. 2018. Language-related domain-specific and domain-general systems in the human brain. Curr Opin Behav Sci. 21:132–137.

Catani M, Jones DK, Ffytche DH. 2005. Perisylvian language networks of the human brain. Ann Neurol Off J Am Neurol Assoc Child Neurol Soc. 57:8–16.

de Almeida RG, Riven L, Manouilidou C, Lungu O, Dwivedi VD, Jarema G, Gillon B. 2016. The neuronal correlates of indeterminate sentence comprehension: an fMRI Study. Front Hum Neurosci. 10:614.

DeLong KA, Troyer M, Kutas M. 2014. Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. Lang Linguist Compass. 8:631–645.

DeLong KA, Urbach TP, Kutas M. 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. Nat Neurosci. 8:1117.

Donhauser PW, Baillet S. 2020. Two distinct neural timescales for predictive speech processing. Neuron. 105:385–393.

Duff MC, Brown-Schmidt S. 2017. Hippocampal contributions to language use and processing. In: The hippocampus from cells to systems. Springer. p. 503–536.

Elman JL. 2011. Lexical knowledge without a lexicon? Ment Lex. 6:1–33.

Federmeier KD, Kutas M. 1999. Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. Cogn Brain Res. 8:373–392.

Federmeier KD, Wlotko EW, Meyer AM. 2008. What's 'Right'in Language Comprehension: Event-Related Potentials Reveal Right Hemisphere Language Capabilities. Lang Linguist Compass. 2:1–17.

Frank SL, Otten LJ, Galli G, Vigliocco G. 2015. The ERP response to the amount of information conveyed by words in sentences. Brain Lang. 140:1–11.

Friederici AD. 2011. The brain basis of language processing: from structure to function. Physiol Rev. 91:1357–1392.

Gajardo-Vidal A, Lorca-Puls DL, Hope TMH, Parker Jones O, Seghier ML, Prejawa S, Crinion JT, Leff AP, Green DW, Price CJ. 2018. How right hemisphere damage after stroke can impair speech comprehension. Brain. 141:3389–3404.

Griffiths TL, Steyvers M. 2004. Finding scientific topics. Proc Natl Acad Sci. 101:5228–5235.

Hagoort P. 2013. MUC (memory, unification, control) and beyond. Front Psychol. 4:416.

Hagoort P, Baggio G, Willems RM. 2009. Semantic unification. In: The cognitive neurosciences, 4th ed. MIT press. p. 819–836.

Hale J. 2001. A probabilistic Earley parser as a psycholinguistic model. In: Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies. Association for Computational Linguistics. p. 1–8.

Hale J. 2006. Uncertainty about the rest of the sentence. Cogn Sci. 30:643–672.

Hickok G, Poeppel D. 2007. The cortical organization of speech processing. Nat Rev Neurosci. 8:393.

Humphries C, Binder JR, Medler DA, Liebenthal E. 2007. Time course of semantic processes during sentence comprehension: an fMRI study. Neuroimage. 36:924–932.

Jung-Beeman M. 2005. Bilateral brain processes for comprehending natural language. Trends Cogn Sci. 9:512–518.

Kamide Y, Altmann GTM, Haywood SL. 2003. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. J Mem Lang. 49:133–156.

Kircher TTJ, Brammer M, Andreu NT, Williams SCR, McGuire PK. 2001. Engagement of right temporal cortex during processing of linguistic context. Neuropsychologia. 39:798–809.

Klimovich-Gray A, Tyler LK, Randall B, Kocagoncu E, Devereux B, Marslen-Wilson WD. 2019. Balancing Prediction and Sensory Input in Speech Comprehension: The Spatiotemporal Dynamics of Word Recognition in Context. J Neurosci. 39:519–527.

Kocagoncu E, Clarke A, Devereux BJ, Tyler LK. 2017. Decoding the cortical dynamics of sound-meaning mapping. J Neurosci. 37:1312–1319.

Korhonen A, Krymolowski Y, Briscoe T. 2006. A Large Subcategorization Lexicon for Natural Language Processing Applications. In: LREC. p. 1015–1020.

Kuperberg GR. 2016. Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. Lang Cogn Neurosci. 31:602–616.

Kuperberg GR, Jaeger TF. 2016. What do we mean by prediction in language comprehension? Lang Cogn Neurosci. 31:32–59.

Kutas M, Federmeier KD. 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu Rev Psychol. 62:621–647.

Lau EF, Namyst A. 2019. fMRI evidence that left posterior temporal cortex contributes to N400 effects of predictability independent of congruity. Brain Lang. 199:104697.

Lau EF, Phillips C, Poeppel D. 2008. A cortical network for semantics:(de) constructing the N400. Nat Rev Neurosci. 9:920.

Levy R. 2008. Expectation-based syntactic comprehension. Cognition. 106:1126–1177.

López JD, Litvak V, Espinosa JJ, Friston K, Barnes GR. 2014. Algorithmic procedures for Bayesian MEG/EEG source reconstruction in SPM. Neuroimage. 84:476–487.

Luke SG, Christianson K. 2016. Limits on lexical prediction during reading. Cogn Psychol. 88:22–60.

Lyu B, Choi HS, Marslen-Wilson WD, Clarke A, Randall B, Tyler LK. 2019. Neural dynamics of semantic composition. Proc Natl Acad Sci. 116:21318–21327.

Maris E, Oostenveld R. 2007. Nonparametric statistical testing of EEG-and MEG-data. J Neurosci Methods. 164:177–190.

Marslen-Wilson WD. 1975. Sentence perception as an interactive parallel process. Science (80-). 189:226–228.

Marslen-Wilson WD. 1987. Functional parallelism in spoken word-recognition. Cognition. 25:71–102.

Marslen-Wilson WD, Tyler LK. 1980. The temporal structure of spoken language understanding. Cognition. 8:1–71.

Marslen-Wilson WD, Tyler LK. 2007. Morphology, language and the brain: the decompositional substrate for language comprehension. Philos Trans R Soc B Biol Sci. 362:823–836.

Marslen-Wilson WD, Tyler LK, Koster C. 1993. Integrative processes in utterance resolution. J Mem Lang. 32:647–666.

Mason RA, Just MA. 2007. Lexical ambiguity in sentence comprehension. Brain Res. 1146:115–127.

Matchin W, Hickok G. 2019. The cortical organization of syntax.

Mattout J, Henson RN, Friston KJ. 2007. Canonical source reconstruction for MEG. Comput Intell Neurosci. 2007.

McClelland JL, McNaughton BL, O'Reilly RC. 1995. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol Rev. 102:419.

Nieuwland MS, Van Berkum JJA. 2006. When peanuts fall in love: N400 evidence for the power of discourse. J Cogn Neurosci. 18:1098–1111.

Price CJ. 2010. The anatomy of language: a review of 100 fMRI studies published in 2009. Ann N Y Acad Sci. 1191:62–88.

Price CJ. 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. Neuroimage. 62:816–847.

Radford A, Narasimhan K, Salimans T, Sutskever I. 2018. Improving language understanding by generative pre-training. URL https//s3-us-west-2 Amaz com/openai-assets/researchcovers/languageunsupervised/language Underst Pap pdf.

Rodd JM, Davis MH, Johnsrude IS. 2005. The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. Cereb Cortex. 15:1261–1269.

Schell M, Zaccarella E, Friederici AD. 2017. Differential cortical contribution of syntax and semantics: An fMRI study on two-word phrasal processing. Cortex. 96:105–120.

Seger CA, Desmond JE, Glover GH, Gabrieli JDE. 2000. Functional magnetic resonance imaging evidence for right-hemisphere involvement in processing unusual semantic relationships. Neuropsychology. 14:361.

Shivde G, Thompson-Schill SL. 2004. Dissociating semantic and phonological maintenance using fMRI. Cogn Affect Behav Neurosci. 4:10–19.

Siri S, Tettamanti M, Cappa SF, Rosa P Della, Saccuman C, Scifo P, Vigliocco G. 2007. The neural substrate of naming events: effects of processing demands but not of grammatical class. Cereb Cortex. 18:171–177.

Smith N, Levy R. 2011. Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. In: Proceedings of the Annual Meeting of the Cognitive Science Society.

St George M, Kutas M, Martinez A, Sereno MI. 1999. Semantic integration in reading: engagement of the right hemisphere during discourse processing. Brain. 122:1317–1325.

Su L, Fonteneau E, Marslen-Wilson W, Kriegeskorte N. 2012. Spatiotemporal searchlight representational similarity analysis in EMEG source space. In: 2012 Second International Workshop on Pattern Recognition in NeuroImaging. IEEE. p. 97–100.

Taulu S, Simola J. 2006. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Phys Med Biol. 51:1759.

Tyler LK, Marslen-Wilson WD. 1977. The on-line effects of semantic context on syntactic processing. J Verbal Learning Verbal Behav. 16:683–692.

Vigliocco G, Vinson DP, Druks J, Barber H, Cappa SF. 2011. Nouns and verbs in the brain: a review of behavioural, electrophysiological, neuropsychological and imaging studies. Neurosci Biobehav Rev. 35:407–426.

Willems RM, Frank SL, Nijhof AD, Hagoort P, Van den Bosch A. 2015. Prediction during natural language comprehension. Cereb Cortex. 26:2506–2516.

Wlotko EW, Federmeier KD. 2007. Finding the right word: Hemispheric asymmetries in the use of sentence context information. Neuropsychologia. 45:3001–3014.

Wright P, Stamatakis EA, Tyler LK. 2012. Differentiating hemispheric contributions to syntax and semantics in patients with left-hemisphere lesions. J Neurosci. 32:8149–8157.

Tables

Table 1: All semantic models used in this study and the epochs in which they were tested against the brain data. The epoch(s) in which each model was tested was chosen specifically to investigate the cascade of incremental predictive processes, 1) emerging with the early activation of the subject noun phrase (SNP) constraint on verbs and on complement nouns (CNs) before the verb is recognised, 2) evolving with a verb being incorporated into the context once the verb is recognised and, 3) facilitating the semantic interpretation once the CN is recognised. The average duration of each word to which each epoch is aligned is indicated by the bracket (mean ± standard deviation).

	Epochs (0-600ms in duration)

	Epoch 1: SN onset (432±142ms)	**Epoch 2: Verb onset (422±111ms)**	**Epoch 3: CN onset (401±115ms)**
Entropy	SNP->verbs SNP->CNs	SNP->verbs SNP->CNs SNP+verb->CNs	SNP+verb->CNs
Mismatch	-	-	SNP+verb<-CN

Captions to Figures

Figure 1: *Overview of the epochs in the experiment in relation to the incremental processing: Epoch 1: Activation of SNP constraint; Epoch 2: Modification of SNP constraint based on the Verb, Epoch 3: Evaluation of SNP+V constraint on CNs. The epochs were each defined relative to an alignment point (AP) such that Epoch 1 is aligned to the SN onset, Epoch 2 is aligned to the verb onset and Epoch 3 is aligned to the complement noun (CN) onset. Each epoch lasted for 600ms which included the average duration of each content word plus one standard deviation. UP = the uniqueness point of a word (the earliest point in time when the word can be fully recognised after removing all of its phonological competitors).*

Figure 2: Design of the experimental stimuli. Each sentence contained a key main verb ("chose") followed by a complement function word ("the" or "to") to vary the complement in terms of the subcategorisation frame preference of a preceding verb. A function word was followed by a noun or a verb that was either consistent with the verb's preferred continuation or less preferred continuation.

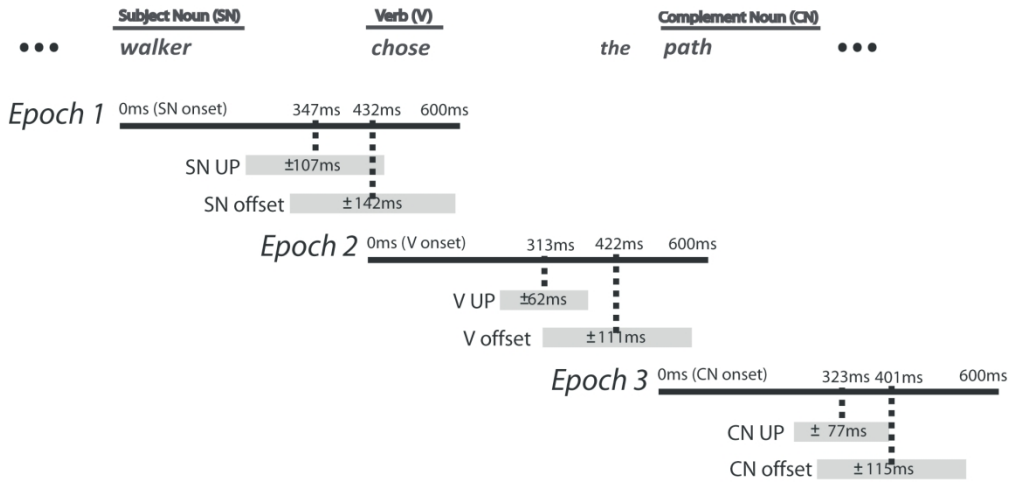
Figure 3: *illustration of reducing entropy in prediction before (left panel) and after (right panel) a verb is incorporated into the SNP context. The topic distributions on the top are the semantic blends of predicted CNs by SNP and SNP+verb respectively. Entropy associated with each of the two distributions is also described. The word boxes below the distributions show a set of preferred words based on the predicted topics.*

Figure 4: A schematic illustration of the searchlight representational similarity analysis of spatiotemporal source-space EMEG data. The bilateral language mask used in this study is surface-rendered onto the brain template in the figure for visualization. Since the source-

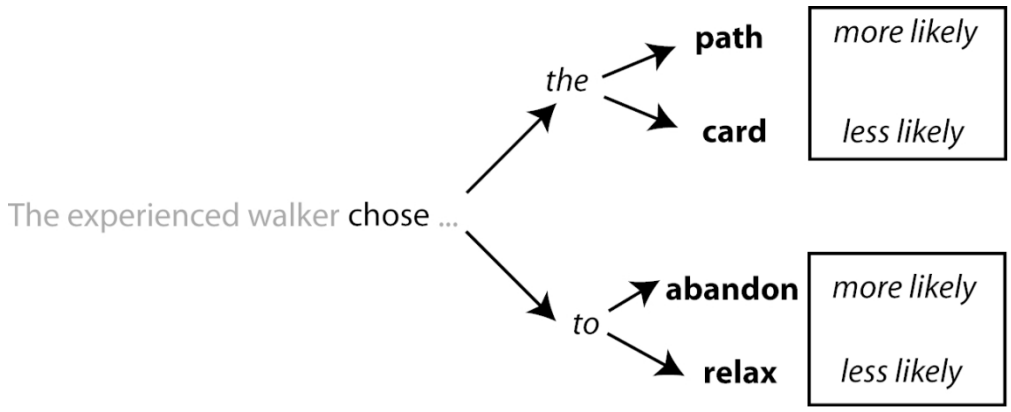
space EMEG data inherently vary across time and space, we calculated the similarity of the spatio-temporal patterns of brain activities for different trials based on measurements within each searchlight sphere with a spatial radius of 10mm and a temporal radius of 30ms. We used 1 – Pearson's correlation between pairs of trials as the distance metric to compute a representational dissimilarity matrix (RDM) for each searchlight, yielding a searchlight map of data RDMs. Each data RDM is then correlated with each model RDM using Spearman's correlation. This Spearman's correlation was computed for each subject and the significance of the correlation at each searchlight location was tested using one-sample t-test (H0: Spearman correlation will be zero). The figure illustrates this process, yielding a time-course of t-values across spatiotemporal searchlights.

Figure 5: *Results of the spatiotemporal searchlight representational similarity analysis with the constraint and mismatch models across three epochs described in Figure 1. Each panel shows the results for different models, corresponding to each subsection in the results. All clusters were corrected by permutation statistics with the cluster-forming threshold (CFT) of p = 0.01 and cluster-wise significance threshold of p = 0.05 (Note that marginally significant clusters with p-values between 0.05 and 0.06 are also reported). A horizontal bar in black indicates the duration of the given cluster. The three alignment points (SN (subject noun), verb and CN (complement noun) onsets) are indicated by long vertical dotted lines. UP stands for "uniqueness point" estimated by the CELEX database and the shaded region in grey around the mean UP reflects ±1 standard deviation from the onset. Similarly, the mean offset of each word is also marked and the region shaded by grey hatch lines around the mean offset reflects ±1 standard deviation from the onset.*

Figure 6: *Vertex-wise peak t-value across three different epochs, summarizing the time-course of t-statistics. Above the surface rendering of each effect, cognitive implications of incremental constraints are illustrated: 1) Activation of broad constraints primarily in right mid-anterior temporal lobe, which additionally engages left temporal and right inferior frontal regions possibly due to the grammatical category and adjacency of a word (CN) being constrained, 2) Developing constraints recruits LIFG-LATL regions that reduce the amount of uncertainty (competition) in the activated constraints and, 3) As a constrained word (CN) is being heard, the specific constraint interacts with the bottom-up input, facilitating its processing in posterior middle temporal and inferior parietal areas.*



Overview of the epochs in the experiment in relation to the incremental processing: Epoch 1: Activation of SNP constraint; Epoch 2: Modification of SNP constraint based on the Verb, Epoch 3: Evaluation of SNP+V constraint on CNs. The epochs were each defined relative to an alignment point (AP) such that Epoch 1 is aligned to the SN onset, Epoch 2 is aligned to the verb onset and Epoch 3 is aligned to the complement noun (CN) onset. Each epoch lasted for 600ms which included the average duration of each content word plus one standard deviation. UP = the uniqueness point of a word (the earliest point in time when the word can be fully recognised after removing all of its phonological competitors).



Design of the experimental stimuli. Each sentence contained a key main verb ("chose") followed by a complement function word ("the" or "to") to vary the complement in terms of the subcategorisation frame preference of a preceding verb. A function word was followed by a noun or a verb that was either consistent with the verb's preferred continuation or less preferred continuation.

108x43mm (300 x 300 DPI)

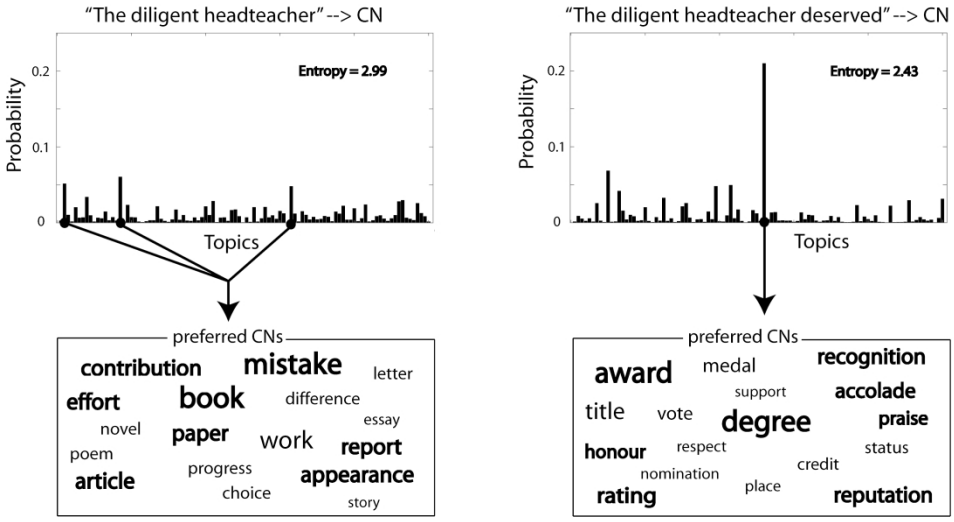
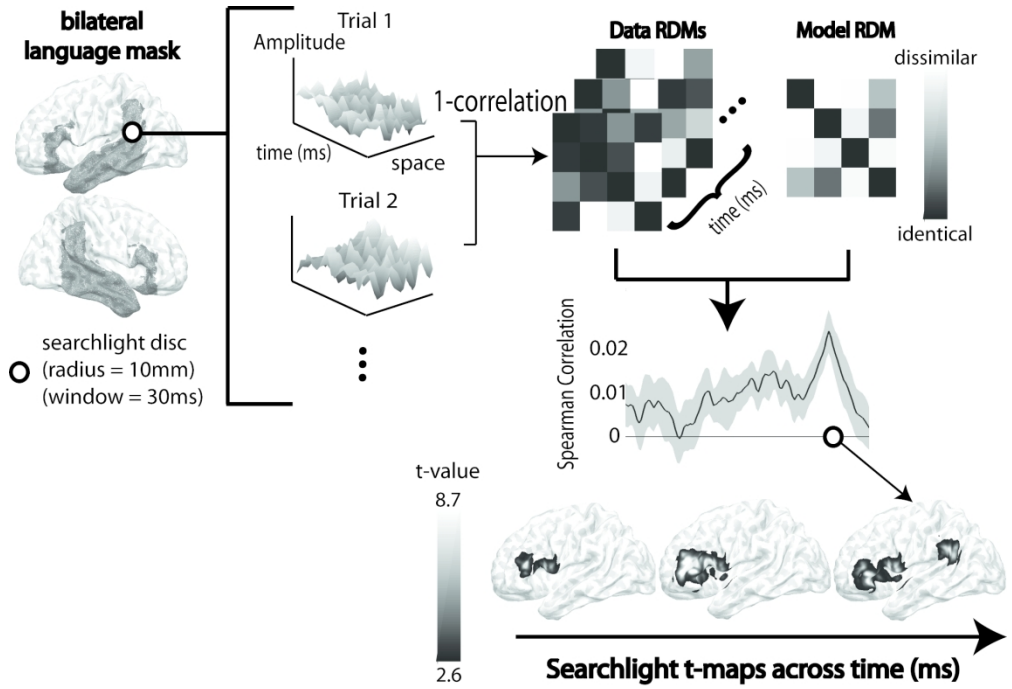
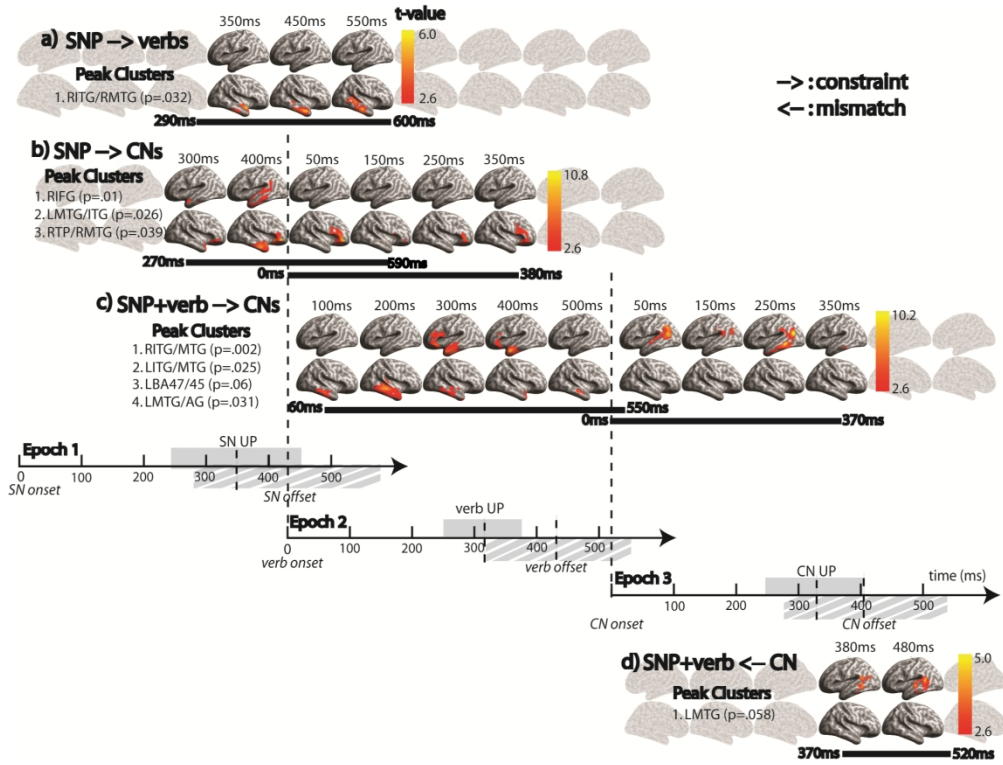


illustration of reducing entropy in prediction before (left panel) and after (right panel) a verb is incorporated into the SNP context. The topic distributions on the top are the semantic blends of predicted CNs by SNP and SNP+verb respectively. Entropy associated with each of the two distributions is also described. The word boxes below the distributions show a set of preferred words based on the predicted topics.

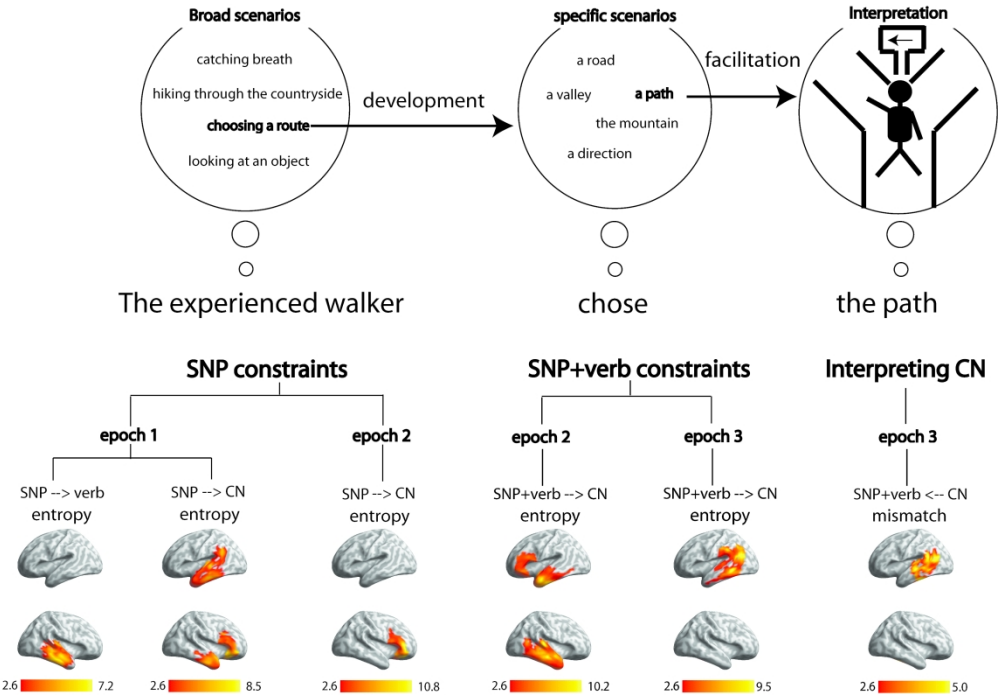
290x148mm (300 x 300 DPI)



A schematic illustration of the searchlight representational similarity analysis of spatiotemporal source-space EMEG data. The bilateral language mask used in this study is surface-rendered onto the brain template in the figure for visualization. Since the source-space EMEG data inherently vary across time and space, we calculated the similarity of the spatio-temporal patterns of brain activities for different trials based on measurements within each searchlight sphere with a spatial radius of 10mm and a temporal radius of 30ms. We used 1 – Pearson's correlation between pairs of trials as the distance metric to compute a representational dissimilarity matrix (RDM) for each searchlight, yielding a searchlight map of data RDMs. Each data RDM is then correlated with each model RDM using Spearman's correlation. This Spearman's correlation was computed for each subject and the significance of the correlation at each searchlight location was tested using one-sample t-test (H0: Spearman correlation will be zero). The figure illustrates this process, yielding a time-course of t-values across spatiotemporal searchlights.



Results of the spatiotemporal searchlight representational similarity analysis with the constraint and mismatch models across three epochs described in Figure 1. Each panel shows the results for different models, corresponding to each subsection in the results. All clusters were corrected by permutation statistics with the cluster-forming threshold (CFT) of p = 0.01 and cluster-wise significance threshold of p = 0.05 (Note that marginally significant clusters with p-values between 0.05 and 0.06 are also reported). A horizontal bar in black indicates the duration of the given cluster. The three alignment points (SN (subject noun), verb and CN (complement noun) onsets) are indicated by long vertical dotted lines. UP stands for "uniqueness point" estimated by the CELEX database and the shaded region in grey around the mean UP reflects ±1 standard deviation from the onset. Similarly, the mean offset of each word is also marked and the region shaded by grey hatch lines around the mean offset reflects ±1 standard deviation from the onset.



Vertex-wise peak t-value across three different epochs, summarizing the time-course of t-statistics. Above the surface rendering of each effect, cognitive implications of incremental constraints are illustrated: 1) Activation of broad constraints primarily in right mid-anterior temporal lobe, which additionally engages left temporal and right inferior frontal regions possibly due to the grammatical category and adjacency of a word (CN) being constrained, 2) Developing constraints recruits LIFG-LATL regions that reduce the amount of uncertainty (competition) in the activated constraints and, 3) As a constrained word (CN) is being heard, the specific constraint interacts with the bottom-up input, facilitating its processing in posterior middle temporal and inferior parietal areas.

Supplementary Information

SI section 1: Topic modelling with Latent Dirichlet Allocation

Modelling the semantics of a word involves taking into account the linguistic context in which it occurs. Unlike feature-based conceptual semantic models (Tyler and Moss 2001), co-occurrence based semantic models naturally reflect the statistical relations among different words based on the fundamental assumption that semantically similar words appear in similar contexts (Harris 1954). Such co-occurrence based semantic models enable semantic contents to be induced from the statistics of large-scale text corpora. Hence, they provide rich distributional content for every word, encoded in the multi-dimensional semantic space whose geometric location and relative distance from the other words define its semantic identity. Such well-defined representational properties allow us to develop reliable models of semantic computations, using quantifiable measures that effectively summarise the representation (or semantic content).

Latent Dirichlet Allocation (LDA) is one of the distributional semantic modelling (DSM) approaches to express the co-occurrence relations in a latent semantic space, assigning every word to one or more latent dimensions in a way that maximises the posterior probability of the model. Just like any other DSM approach, it is built upon the distributional hypothesis which claims that any words that occur in similar (linguistic) context are semantically similar (Harris 1954). Its distinct quality is in its formulation in a probabilistic (Bayesian) framework that takes the advantage of using a Dirichlet prior, the conjugate distribution of the multinomial likelihood. In this way, it is possible to marginalize the parameter(s) and express the Dirichlet posterior in terms of the known variable (i.e. observed samples) and the hyper-parameter(s).

The model training involves iterative updating of two probability distributions, known as target-topic (e.g. $P(target\_word \mid topic)$) and topic-context (e.g. $P(topic|context\_word)$) distributions:

$$P(target\_word \mid context\_word) = \sum_{topic} P(target\_word \mid topic) P(topic|context\_word)$$

Each of these distributions was parameterized by separate multinomial variables (e.g. $\emptyset$ and θ) that specify a distribution either over target words for a given topic (target-topic), or over topics for a given context (topic-context). Then, using collapsed Gibbs sampler, we computed the maximum a posterior (MAP) estimate of each of these parameters as the following:

$$E_{pos}[\emptyset_{j,w_i}] = P(w_i \mid z_i = j, z_{-i}, w_{-i}) = \frac{f^{(w_i)}_{-i,j} + \beta}{f_{-i,j} + |W|\beta}$$

$$E_{pos}[\theta_{c_i,j}] = P(z_i = j \mid z_{-i}, c_i) = \frac{f^{(c_i)}_{-i,j} + \alpha_j}{\sum_j f^{(c_i)}_{-i,j} + \alpha_j}$$

where i and j are indices to observed samples and latent dimensions (topics) respectively, w_i, c_i and z_i are the $target\_word$, $context\_word$ and $topic$ at the i^{th} (current) observation, $-i$ represents all observations other than the i^{th} observation, β and α_j are the symmetric and asymmetric hyper-parameters (Wallach et al. 2009) associated with $\emptyset$ and θ respectively, $|W|$ is the total number of words in the vocabulary and f represents the frequency count such that $f^{(w_i)}_{-i,j}$ is the frequency of a word at the current observation i associated with the topic j after taking out a topic assignment at i. The hyper-parameters were optimised in a way that maximises the model evidence using the fixed-point iteration scheme (Minka 2000).

For the actual model training, the two hyper-parameters were initially randomized but were jointly updated for every observation in the training dataset by re-sampling the topic after taking out the randomly assigned topic and computing the leave-one-out probability distributions. This training approach is known as collapsed Gibbs sampler (a variant of Gibbs sampling that involves marginalization of the multinomial parameters), one of the well-known Markov chain Monte Carlo (MCMC) methods which obtain a sample by observing the chain (whose desired distribution is same as its equilibrium distribution) after a number of training steps. Here, each chain refers to a stochastic model at a given training step whose probability distributions are computed based solely on the previous step such that the future state of the system is conditionally independent to its past states given its present state. Not surprisingly, the random sample (or a state of the model at each step) from this MCMC method is inherently auto-correlated. Therefore, we took the distributions from three sampling states which were 50 training steps apart from each other (to maximize the degree of independence between sampling states) after the burn-in period of 200 steps (for model stabilization). The total number of topics was set to 100 ($|Z| = 100$). All these parameters were set to be consistent with (Ó Séaghdha and Korhonen 2014). The distributions from these samples were averaged and the averaged topic-context distribution was used as a model of semantic representation. See Figure S1 for an illustration of the explanatory values of our topic model.

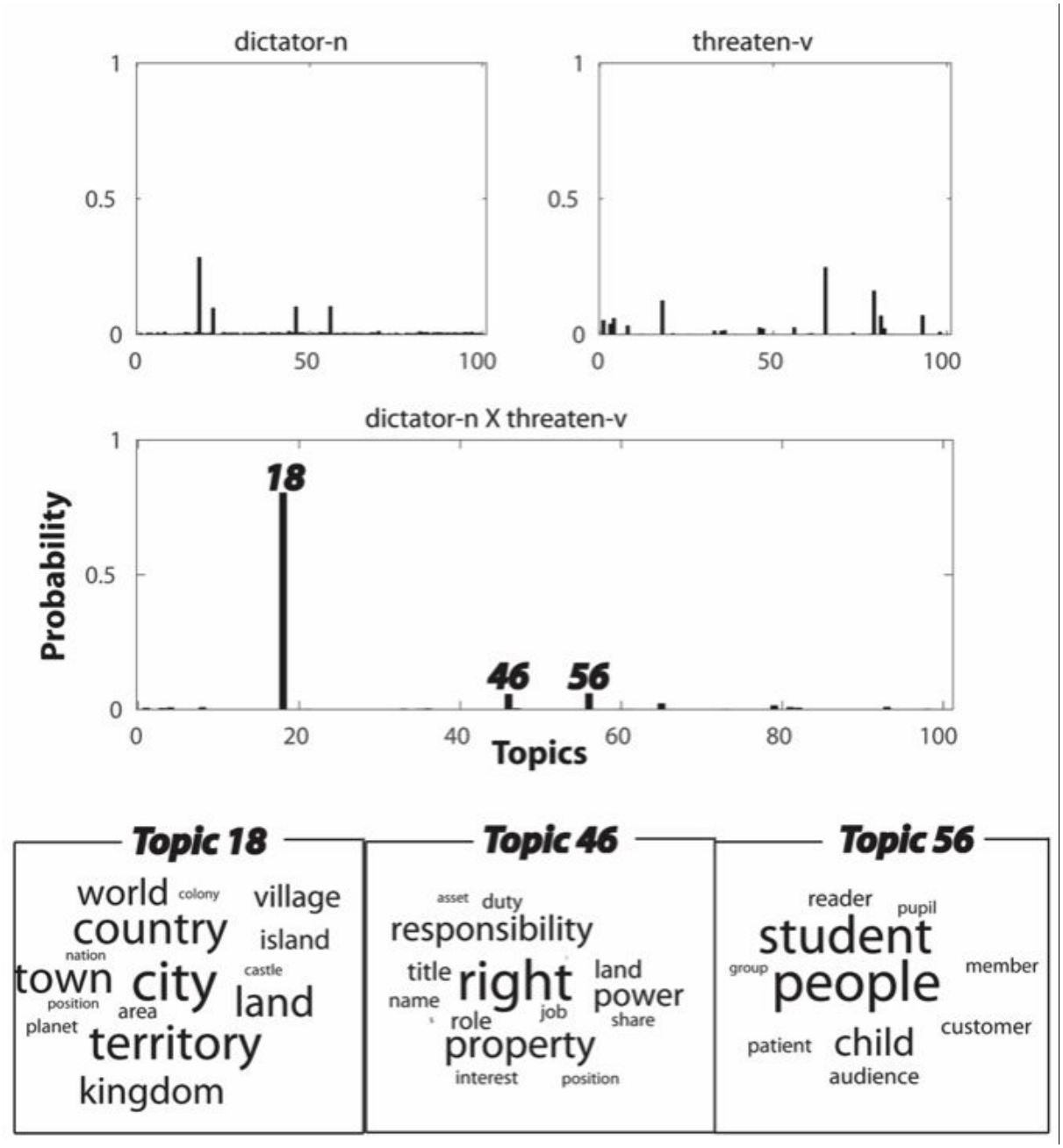


Figure S1: *A visual illustration of our SNV-CN topic model. In the top row, the predictive activations of "dictator" (SN) and "threaten" (verb) are depicted by our topic-context distributions. In the middle row, we model the consistency in topic preference between "dictator" and "threaten" using an element-wise multiplication (.*) as a combinatorial operator (re-normalised for visualisation). Lastly, in the bottom row, a list of complement nouns (CNs) preferred by each of the top three consistent topics is visualised (The size of each CN in the list reflects the target-topic probability from our topic model).*

References

Harris ZS. 1954. Distributional structure. Word. 10:146–162.

Minka T. 2000. Estimating a Dirichlet distribution.

Ó Séaghdha D, Korhonen A. 2014. Probabilistic distributional semantics with latent variable models. Comput Linguist. 40:587–631.

Tyler LK, Moss HE. 2001. Towards a distributed account of conceptual knowledge. Trends Cogn Sci. 5:244–252.

Wallach HM, Mimno DM, McCallum A. 2009. Rethinking LDA: Why priors matter. In: Advances in neural information processing systems. p. 1973–1981.

Supplementary Information

SI section 2: Effect size analysis

In this section, we report the results of our effect size analysis, showing that all of our results presented in the main text are reliable. We carried out an effect size analysis where we calculated an effect size map (i.e. Cohen's D) and a power map based on $P(t_{obs} - t_{crit} > 0)$ where t_{obs} is the observed t-value for every data-point (searchlight) across space and time and t_{crit} is the critical t-value determined by the false positive rate (alpha) = 0.05. For statistical summary, we computed an average effect size (D) and power across all searchlights within a cluster reported in the main text (Figure 5).

Overall, our results clearly showed a strong effect size (>0.8) and a reliable statistical power (~90%) for all of the clusters reported in this study (see Figure S2) (compare Figure S2 with Figure 5 in the main text). The mismatch analysis at Epoch 3 based on 128 items (stimuli) had the lowest, yet strong effect size and power (see panel (d) in Figure S2). This is likely because we used a stringent cluster-forming threshold (CFT: p=.01 instead of p=.05) which compensates for the modest sample size (N=13). Based on these findings, we claim that the reported effects are strong and reliable.

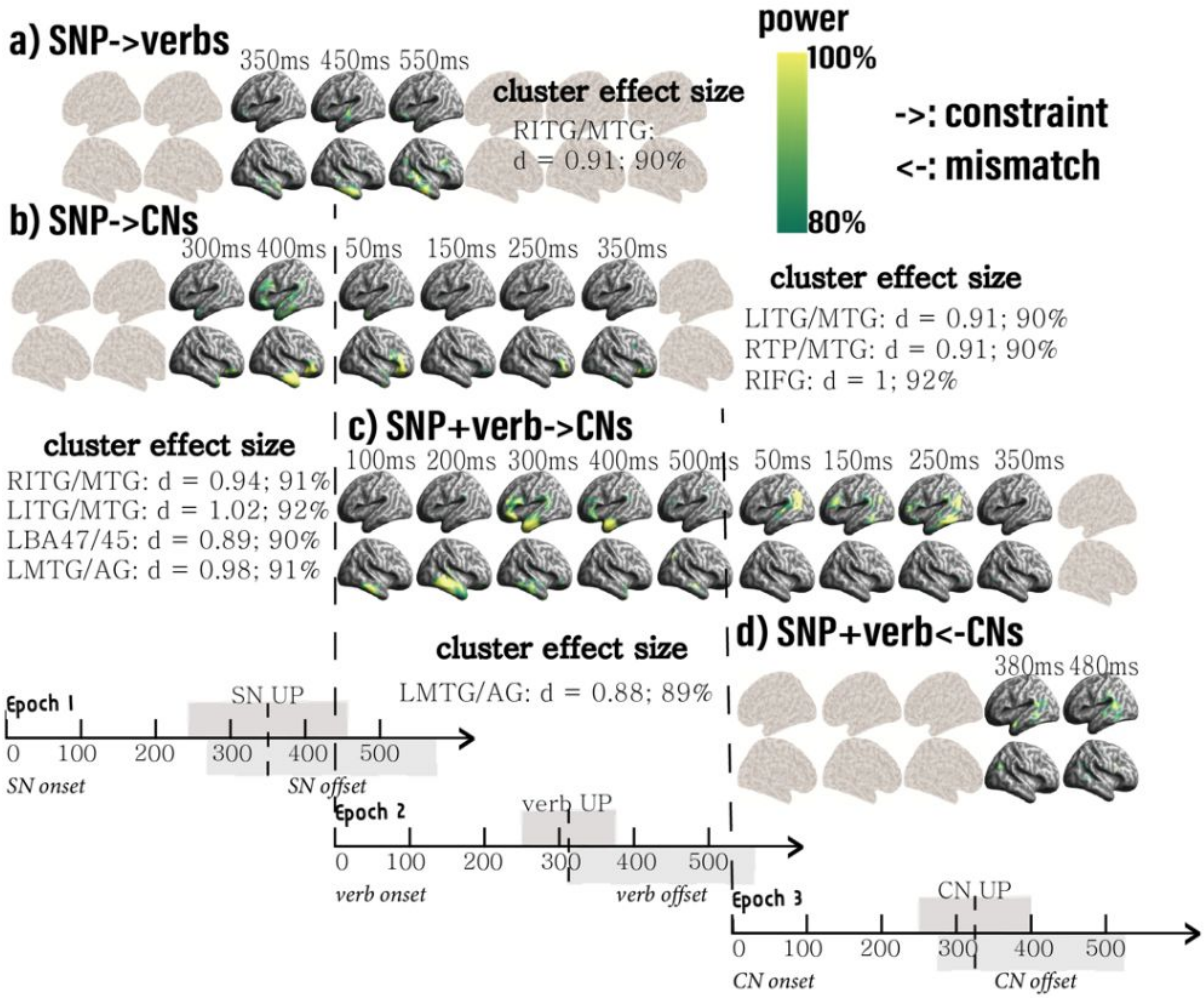


Figure S2: *Results of the power analysis with the current sample size N = 13. An additional power analysis was conducted for all of the results presented in Figure 5 in the main text at the same epochs. For visualization, we rendered any vertices above 80% power threshold but the cluster effect size in each of the four panels shows the effect size statistics (Cohen's d) and the power of our t-statistics averaged over all searchlights within each cluster reported in the main text.*

Supplementary Information

SI section 3: Entropy reduction (evolving constraint) analysis

If the anteroventral network (LIFG-LATL) subserves the integrative process of unifying the verb into the SNP (see (c) in Results), they essentially will lead to a reduction in uncertainty by pruning any irrelevant topics based on the verb. In order to demonstrate this uncertainty reduction in our stimuli, we statistically compared the entropy values of the semantic blends before and after the verb by running a paired-sample t-test (i.e. we statistically compared the two entropy models which were described in 3.4-2) and 3.4-3) in Methods and reported in b) and c) in Results). Consistent with our prediction, we found significant reduction in entropy after adding a verb to the context (t(45)=7.4, p<.001). See Figure S3-1 for further descriptive statistics.

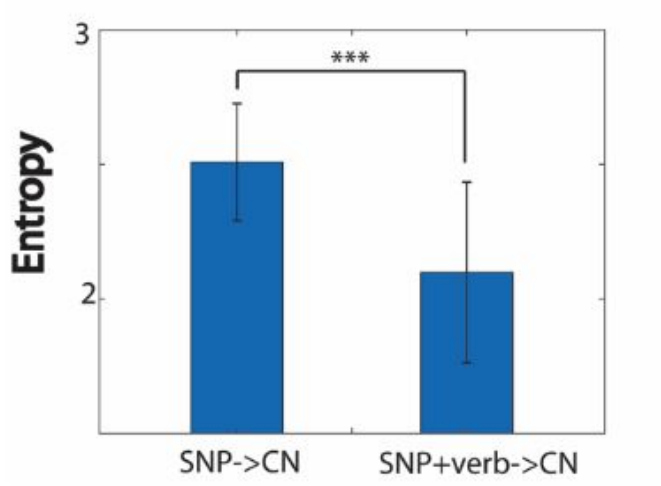


Figure S3-1: *Description of mean and standard deviation of entropy before and after integrating a verb into the SNP context. *** indicates significant difference (p<.001).*

Following this analysis, we further constructed the entropy change model which captures how specific the constraint has become by integrating a verb into the context. It was simply calculated as the amount of reduction in entropy after integrating a verb into the SNP context:

$$H[P(CN\ topic \mid SNP)] - H[P(CN\ topic \mid SNP + verb)]$$

In order to investigate how semantic constraint is neurally evolved, this model was tested in Epoch 2 and 3 using the exactly same ssRSA analysis pipeline as described in the Methods section 4.
To our surprise, we found a significant cluster not in the LIFG-LATL regions but in the RH mid-anterior temporal regions from 180ms to 510ms in Epoch 2 (p=.016; see panel (a) in Figure S3-2).

This may suggest the role of the RH mid-anterior temporal areas in both early construction and computational development of the context-based semantic constraint during natural speech comprehension. However, it does not fully explain the role of LIFG-LATL areas during incremental predictive processing.

Based on our previous findings (Lyu et al. 2019) which showed the verb-CN interaction effect in LIFG (i.e. integrated semantic representation above and beyond the semantic representation of individual words), we hypothesized that combinatorial processing of lexically-based constraints primarily recruits LIFG, projecting to the LH temporal cortex via ventral pathways (e.g. uncinate fasciculus and extreme capsule). When constraining CNs, the preceding SN and verb are the two most important words, setting up a scene for an upcoming CN. Hence, it is often very difficult to fully separate the context-based from words-based constraints. In this study, we tested an additional entropy change model calculated as: $H[P(CN\ topic \mid SN)]$ - $H[P(CN\ topic \mid SN + verb)]$. $P(CN\ topic \mid SN)$ was taken directly from our second SNV-CN topic model (see 3.2 in Methods) and $P(CN\ topic \mid SN + verb)$ was calculated as an element-wise multiplication between the SN-topic and verb-topic distributions:

$$P(CN\ topic \mid SN + verb) = P(CN\ topic \mid SN). * P(CN\ topic|verb)$$

which was normalized into a probability scale.

Testing this word-based entropy reduction model yielded a significant effect in LIFG-LATL areas, but at the later stage while hearing the CN approximately until its offset (L-BA45/47: p=.011 and LITG/MTG: p=.032; see panel (b) in Figure S3-2). This likely reflects the ongoing fine-tuning of the constraint while processing the bottom-up input. In addition, there was a hint of such computation earlier in LIFG from 230ms to 420ms in Epoch 2 which did not reach the significance threshold after the multiple comparisons correction (L-BA45: p=.09).

To further clarify if this lexically based entropy change model accounts for the LIFG-LATL clusters of the full-context SNP+verb constraint on CNs, we carried out an additional statistical analysis where we added this entropy change model as a convariate and partialled out when correlating the SNP+verb constraint model RDM with the searchlight data RDMs. Then, the output correlation map was 1) statistically tested against zero in the same way as described in the Methods section 8, and 2) statistically compared with the original correlation map without adding the entropy change model as a covariate. For this purpose, we used a one-tailed paired-sample t-test, testing against a null hypothesis that there is no difference between the two correlation maps, paired across subjects. Only for this contrast analysis, we specified the contrast window from 230ms to 460ms (the time window during which significant clusters of SNP+verb constraint on CNs emerged in LIFG-LATL areas

(see panel (c) in Figure 5)). These two statistical analyses were conducted after the verb-onset in Epoch 2.

First, we found two clusters of the SNP+verb constraint on CNs in the bilateral temporal regions from 50ms to 460ms after the verb-onset when lexically-based entropy change was partialled out (p=.002 in RITG/MTG and p=.047 in LITG/MTG; see panel (a) in Figure S3-3). Second, we showed that the SNP+verb constraint model fit is significantly reduced in LIFG from 240ms to 450ms after the verb-onset (p=.031 in L-BA45; see panel (b) in Figure S3-3). Combining these results, we suggest that the process of integrating semantic constraint from multiple lexical sources recruits L-BA45/47 which interactively develops the context-level constraint to become more specific. Although this is highly consistent with previous findings and neurobiological accounts of speech comprehension (Hagoort 2013; Kocagoncu et al. 2017; Lyu et al. 2019), future studies that investigate the spatiotemporal neural dynamics during incremental sentence/discourse comprehension will further establish the role of multiple brain regions in the language network beyond lexical processing.

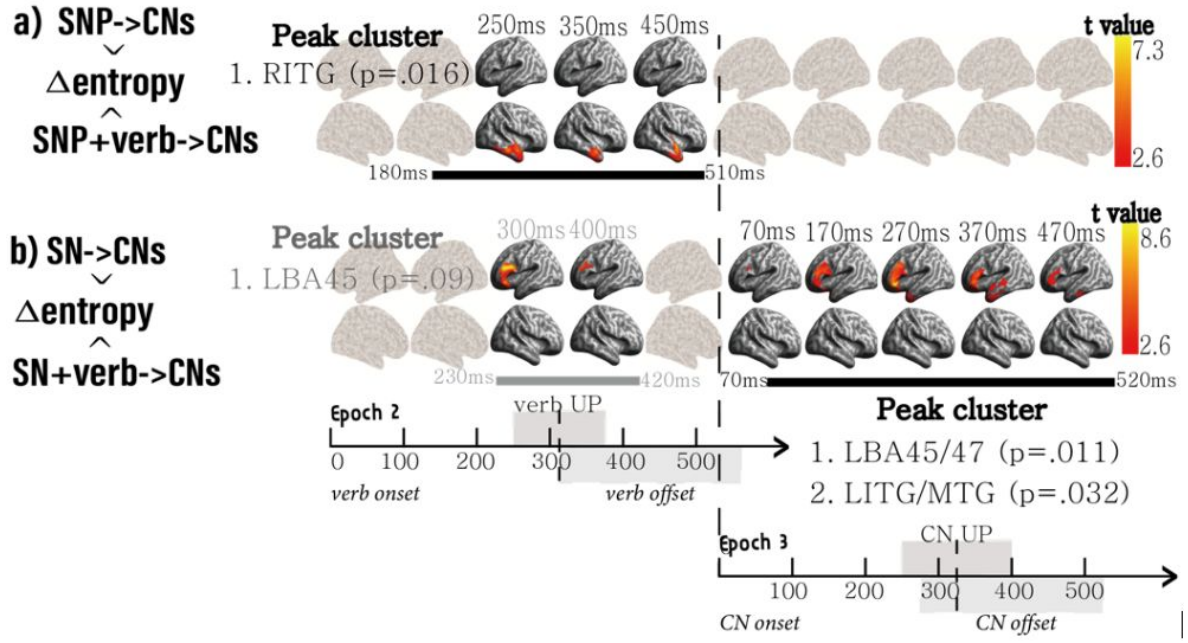


Figure S3-2*: Results of an additional analysis with two different entropy change models. A panel a) at the top shows the clusters found by entropy change of the full-context whereas a panel b) at the bottom shows the entropy change effects captured by the lexical properties of a subject and a verb. The cluster presented in panel b) in grey was not significant after the multiple comparisons correction (p=.09).*

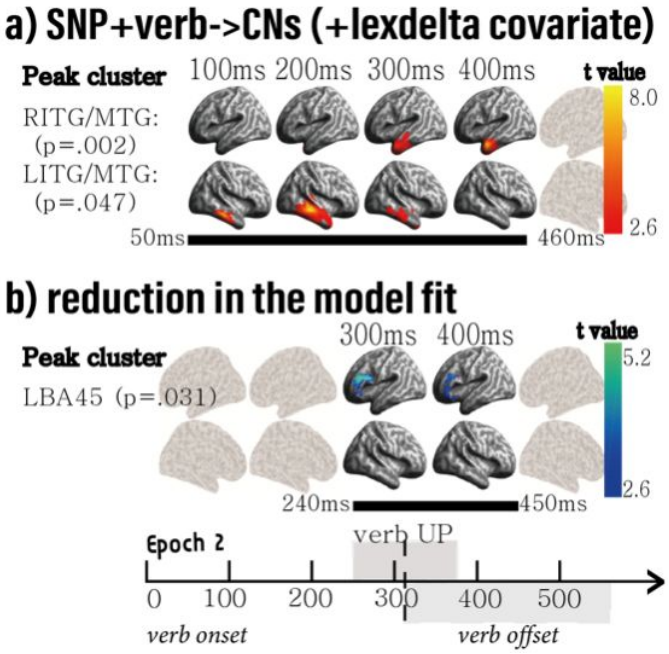


Figure S3-3: *Results of an additional partial correlation analysis by partialling out the covariate of interest when testing the SNP+verb constraint model shown in the panel c) in Figure 5 of the main text. The covariate of interest (denoted as lexdelta in this figure) is the entropy change model based on the lexical properties of a subject (SN) and a verb presented in the panel b) of Figure S3-2). A panel a) shows all clusters found by the SNP+verb constraint in this analysis and a panel b) shows a cluster that was explained away by the covariate.*

References

Hagoort P. 2013. MUC (memory, unification, control) and beyond. Front Psychol. 4:416.

Kocagoncu E, Clarke A, Devereux BJ, Tyler LK. 2017. Decoding the cortical dynamics of sound-meaning mapping. J Neurosci. 37:1312–1319.

Lyu B, Choi HS, Marslen-Wilson WD, Clarke A, Randall B, Tyler LK. 2019. Neural dynamics of semantic composition. Proc Natl Acad Sci. 116:21318–21327.

Supplementary Information

SI section 4: Cross-covariance decomposition analysis

The finding that both SNP constraints on verbs and on CNs elicit similar activation in RH temporal regions with similar timing raises the possibility that these constraints are initially activated by the same event representation based on the SNP (see (a) and (b) in Results). If this is the case, we would expect to see some degree of overlap in the predicted semantic space between verbs and CNs; for example, a set of verbs constrained by the preceding SNP would share a similar pattern of topic preferences across stimuli with a set of constrained CNs. To test this hypothesis, we ran an additional cross-covariance decomposition analysis where we projected the blended representations of predicted verbs (see 3.3a in Methods 3.3a) and CNs (see 3.3b in Methods) to a latent subspace that maximally explains the cross-covariance structure between them.

Analysing cross-covariance is one of the simple approaches to quantify the relationship between two multivariate datasets. In this study, we statistically tested if the predicted verbs and CNs based on the SNP context (i.e. blends) share some overlapping semantic space. To do this, we decomposed the cross-covariance between these blends which were centred to have a mean of zero and projected them onto a common subspace that maximally explains the co-variability in the data across topics. However, as our datasets (blends) consist of 50 unique SNPs and 100 topics (features), this approach is highly susceptible to overfitting. To prevent this issue, we carried out 5-fold cross-validation with 100 random partitions splitting the items (trials) into training and test sets for both blends. The training set was used to compute the loading vectors onto which the test set was projected, using singular value decomposition (SVD) of a cross-covariance matrix between the two blends (i.e. The loading vectors were the first left and right singular vectors). Then, we computed a correlation coefficient between the two blends in the projected subspace. Lastly, for statistical testing, we obtained a null-distribution by randomly permuting the items in one of the two blends. This random permutation was repeated 1,000 times and the output correlation coefficient under the null was saved for each iteration.

The result showed a statistically significant correlation between the predicted verbs and CNs once they were projected onto the first canonical basis that maximally explains the cross-covariance ($r=0.58$; $p<.001$; see Figure S4). This additional result confirms a largely overlapping semantic space between predicted verbs and CNs, plausibly reflecting the event representation generated by the SNP context.

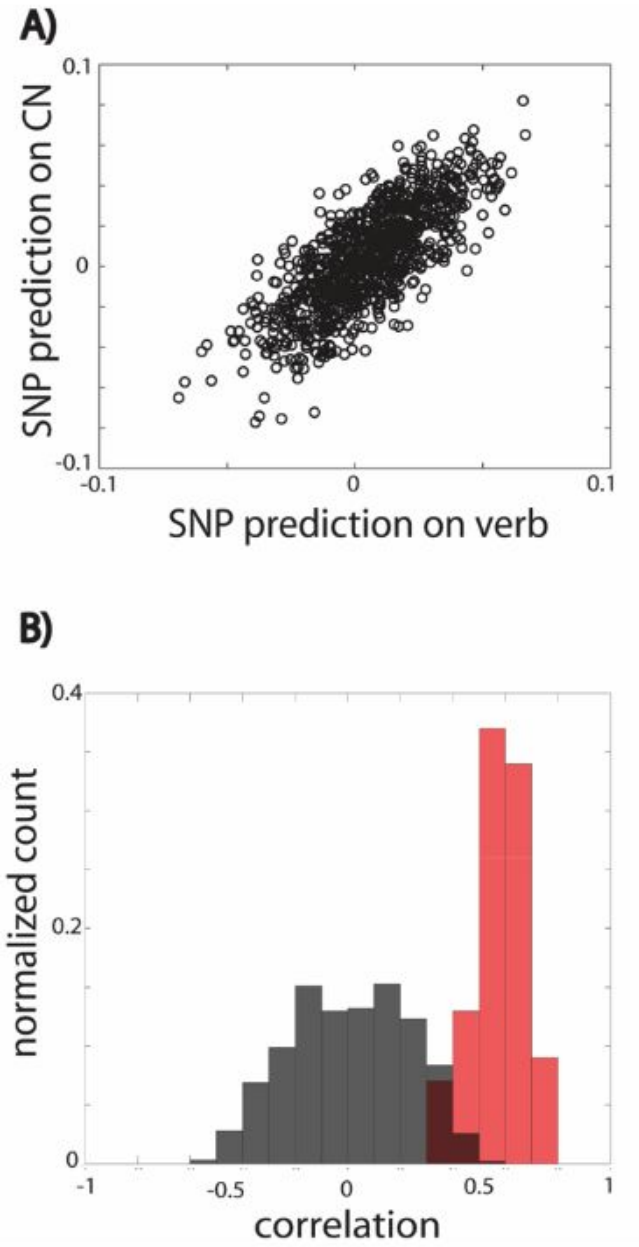


Figure S4: *Panel A) illustrates the relationship between the predicted verbs and CNs (blends) after projecting onto the one-dimensional subspace that maximally explains the cross-covariance between them. The plot concatenated the averaged 10 data-points from the testing set across 100 partitions. Panel B) shows two different histograms: 1) a null-distribution across 1,000 permutations in grey and 2) the actual correlations across 100 partitions in red. We visualised the variability across 100 random partitions to highlight the reliability of our result and the mean correlation value was 0.58.*

Supplementary Information

SI section 5: Modelling event-level semantic constraint

In the main text, we demonstrated how context semantically constrains upcoming words. However, understanding speech not only entails such lexical processing, but it also requires building an event representation depicting the underlying message that the speaker conveys. In order to further investigate the neural underpinnings of early event-level prediction, we constructed a model that captures a joint constraint of the SNP context on verb and CN.

To do this, we blended the topic distributions from the second SNV-CN topic model (see 3.2 in Methods) based on the behavioural responses from the first sentence completion study (see 3.1 in Methods). We first calculated a probabilistic weight $P(verb + CN|SNP)$ by counting each unique continuation response of verb and CN after hearing SNP. We also computed the conditional distributions $P(CN\ topic|verb + CN)$ for each unique continuation response, using an element-wise multiplication between the two vectors $P(CN\ topic|verb)$ and $P(CN\ topic|CN)$. Then, we blended these joint distributions as described in the Methods section 3.3:

$$P(CN\ topic \mid SNP) = \sum_{verb+CN} P(CN\ topic \mid verb + CN)P(verb + CN|SNP)$$

Although this event blend characterizes the same CN topic distribution as the SNP's non-adjacent constraint on CNs, this particular formulation jointly depicts the semantic constraint of the SNP on verbs and CNs which are optimally expressed in the common latent dimensions (named as $CN\ topic$ throughout this paper). Then, we calculated the entropy of this event blend as we did for that of SNP constraints on verb and CN separately and tested this model in Epoch 1 and 2 in the exactly same ssRSA analysis pipeline as described in Methods section 4. We obtained the following results (see Figure S5):

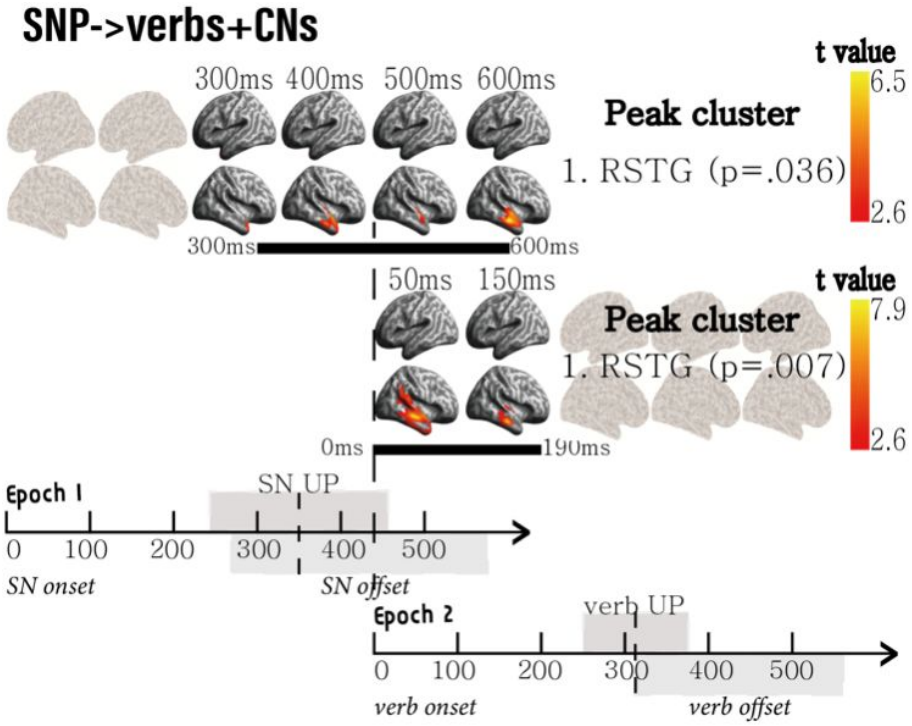


Figure S5: *Results of an additional analysis with the event constraint model (i.e. SNP constraint on the integrated semantics of verbs and CNs). Epoch 1: A significant cluster from 300ms to 600ms after the SN onset, peaking the anterior portion of STG in RH (p=.036). Epoch 2: A significant cluster from 0ms to 190ms after the verb onset peaking at the RH STG (p=.007). See Figure 5 in the main text for more illustrative details of the figure.*

Combining these findings, we can confirm that the overall event-level constraint (i.e. The SNP constraint on combined verbs and CNs) was initially activated in the RH mid-anterior temporal regions approximately as the SN was recognised. This pattern of results is largely consistent with the early SNP constraints on each individual upcoming word including verbs (see panel (a) in Figure 5) and CNs (see panel (b) in Figure 5). However, unlike the SNP constraints on each individual upcoming word, this overall event constraint appears in Epoch 2 in RH mid-STG/MTG areas approximately until 200ms into the verb. This pattern of results supports our interpretation that the early SNP constraint in RH mid-anterior temporal regions not only constrains each individual upcoming word but it also constrains a combination of upcoming words (scenarios). Consistent with our expectation, the event constraint of SNP disappeared before the SNP constraint on CNs represented in RIFG which might reflect maintenance until the verb is integrated into the SNP context.